Human Metacognition Inspired Learning Algorithms in Neural Networks and Particle Swarm Optimization

Asst. Prof. Sundaram Suresh

Nanyang Technological University, Singapore

Email : ssundaram@ntu.edu.sg

Webpage: www.ntu.edu.sg/home/ssundaram

IEEE SSCI 2015 Tutorial

Cape Town, South Africa

Organization

- A brief review on human learning from cognitive psychology.
- Part I – Neural Networks
  - Self-regulated learning with meta-cognition in machine learning
  - Meta-cognitive neural networks and its self-regulatory learning algorithms
    - PBL-McRBFN classifier
  - Benchmark evaluation and comparisons
  - Applications in Medical informatics
- Part II – Particle Swarm Optimization
  - Self-regulated particle swarm optimization
  - Dynamic mentoring based particle swarm optimization
  - Real-world applications
- Future directions

Motivation

- Self-regulated learning is the best learning strategy [1],[2],[3]

- Student control their learning process not teacher
  - Set their own goal (**plan**)
  - Identify what to learn and choose material: video lecture, book, (**monitor**)
  - Get feedback on their understanding (**manage**)

Class Room Learning

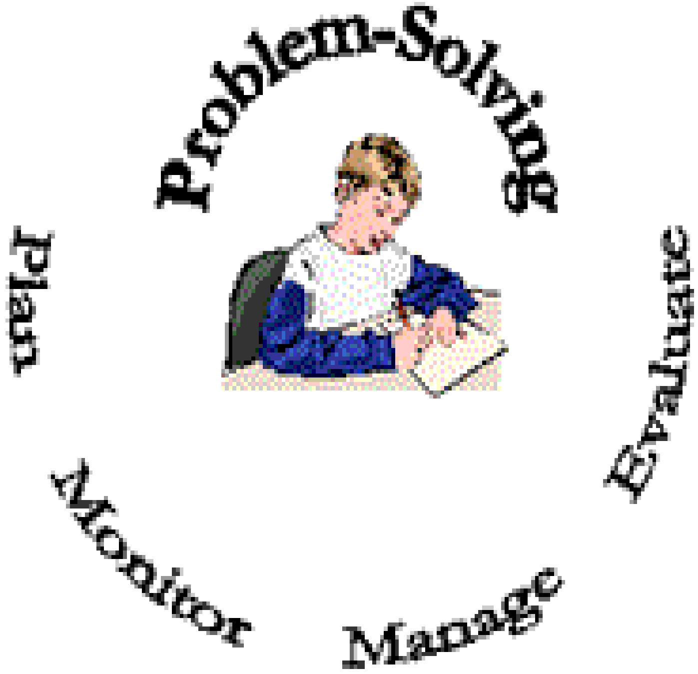
- Human meta-cognition controls the learning process

[1] Wenden, A. L. (1998). Metacognitive knowledge and language learning. *Applied Linguistics*, 19(4), 515–537.
[2] Rivers, W. P. (2001). Autonomy at all costs: An ethnography of meta-cognitive self-assessment and self-management among experienced language learners. *Modern Language Journal*, 85(2), 279–290.
[3] Isaacson, R., & Fujita, F. (2006). Metacognitive knowledge monitoring and self-regulated learning: Academic success and reflections on learning. *Journal of the Scholarship of Teaching and Learning*, 6(1), 39–55.

Definition of Metacognition

- “The awareness and knowledge of one’s mental processes such that one can monitor, regulate and direct them to a desired goal”
  - As defined by J.H Flavell (1976)

What is self- regulation?



- For effective learning, learners employ self- regulation [2],[3].

[2] Rivers, W. P. (2001). Autonomy at all costs: An ethnography of meta-cognitive self-assessment and self-management among experienced language learners. *Modern Language Journal*, 85(2), 279–290.
[3] Isaacson, R., & Fujita, F. (2006). Metacognitive knowledge monitoring and self-regulated learning: Academic success and reflections on learning. *Journal of the Scholarship of Teaching and Learning*, 6(1), 39–55.

Definition

- Self-regulation
  - An active constructive process whereby learners set goals, monitor, regulate, and control their cognitive and metacognitive process in the service of their goals
  - Provide role in collaborative learning

Why Meta-cognition is important?

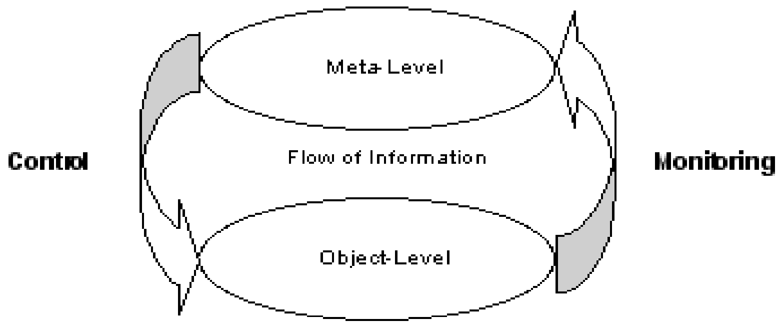
- Learning
  - How-to-learn?
  - What-to-learn?
  - When-to-learn?
- Helps to promote "Deep Learning"
- Assessment for learning
  - Active role in assessing his/her own learning
  - Encourage to take responsibilities
  - Provide awareness

Models of Metacognition

Nelson and Naren Model

- Cognitive component
  - Represent the knowledge
- Metacognitive component
  - Represent dynamic model of the cognitive component
- Signals
  - Control
    - Change the state of cognitive component or cognitive component itself
    - Initiate, or terminate or continue
  - Monitory
    - Inform about cognition

Nelson and Narens Model

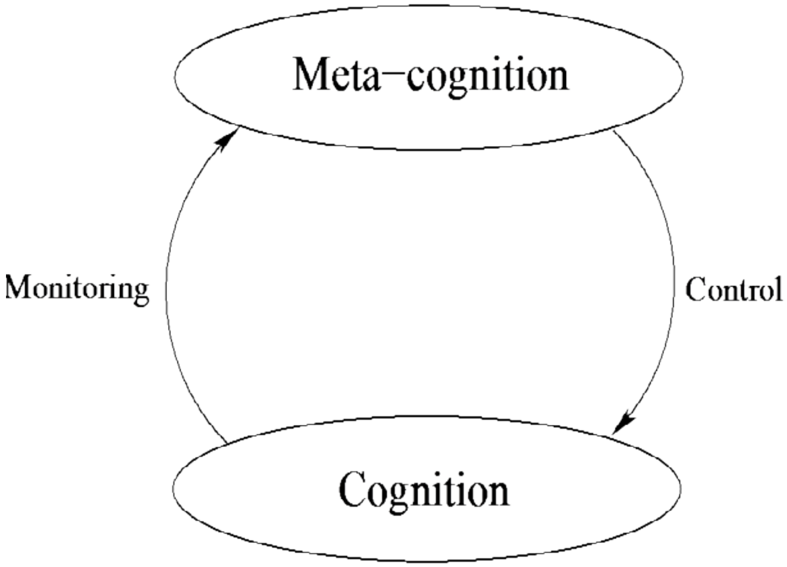


Nelson T. O and Narens L, Metamemory: A theoretical framework and new findings, Psychology and Learning Motivation, 26, pp. 125-173, 1990.

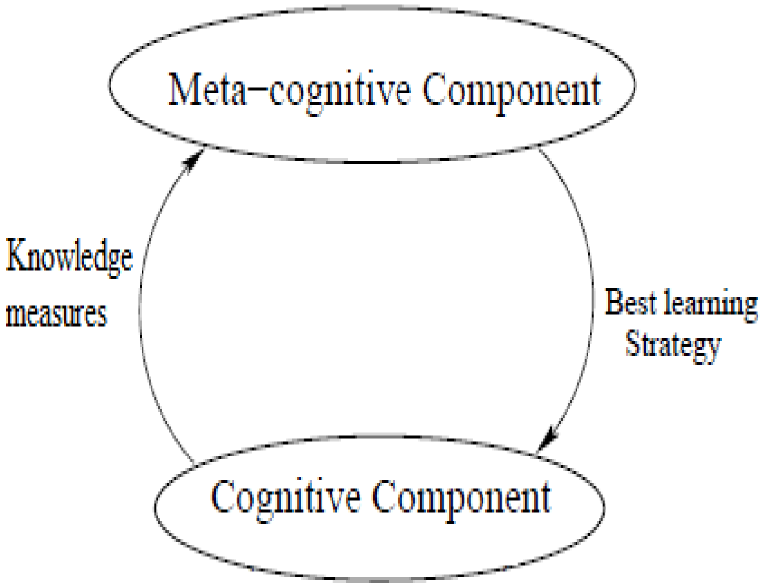
Part I
How metacognition is incorporated in neural networks?

Metacognitive network

Nelson Naren's Model



Metacognitive network



Meta-cognitive network

Cognitive component

- Representation of knowledge
  - To be learnt from the sample stream
  - Unknown
    - Suitable structure and its parameters
- Choice of knowledge representation
  - Neural network : RBFN
  - Neuro-Fuzzy
  - Complex-valued neural network
  - etc..

Meta-Cognitive component

- Learning about learning
  - Decides
    - What-to-learn
      - Proper choice of samples from stream based on current state of knowledge
    - When-to-learn
      - Appropriate usage of sample in right interval
    - How-to-learn
      - Structure modification
      - Parameter learning

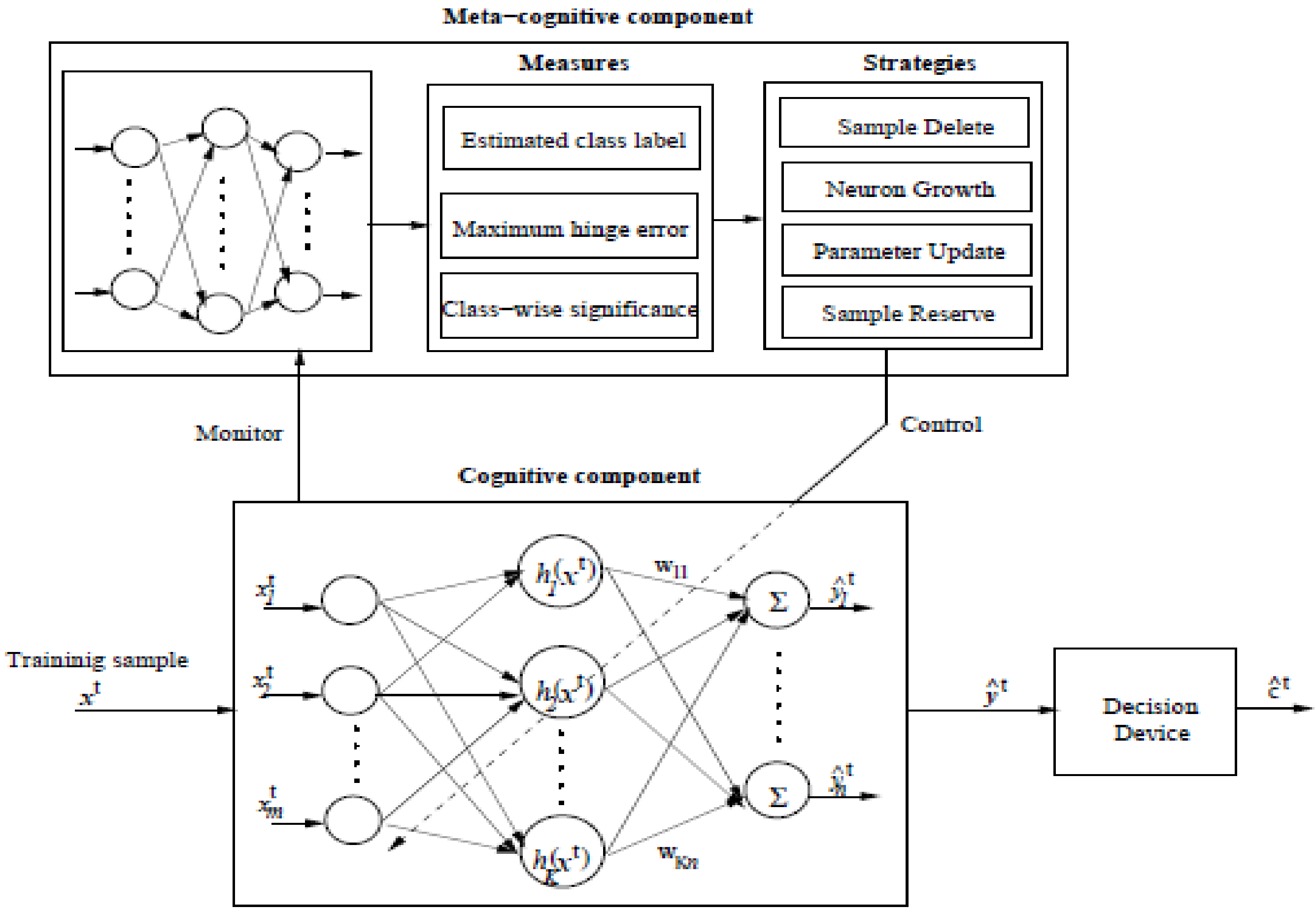
Current state of metacognitive networks

- Neural Network
  - G. Sateesh Babu, S. Suresh, Sequential projection based metacognitive learning in a Radial basis function network for classification problems, IEEE Trans, on Neural Networks and Learning Systems, 24(2), pp. 194-206, 2013.
  - G. Sateesh Babu, and S. Suresh, Meta-cognitive RBF networks and Its Projection based Learning Algorithm for Classification problems, Applied Soft Computing, 13(1), pp. 654-666, 2013.
  - G. Sateesh Babu, and S. Suresh, Meta-cognitive neural network for classification problems in a sequential learning framework, Neurocomputing, 81(1), pp. 86-96, 2012.
- Neuro-Fuzzy systems
  - K. Subramanian, S. Suresh, N. Sundararajan, "A meta-cognitive neuro-fuzzy inference system (McFIS) for sequential classification problems, IEEE Trans. on Fuzzy Systems, 2013
  - K. Subraminan, and S. Suresh, A meta-cognitive sequential learning algorithm for neuro-fuzzy inference system, Applied soft computing, 12(11), 36703-3614, 2012.
- Complex-valued neural network
  - R. Savitha, S. Suresh, and N. Sundararajan, Metacognitive learning algorithm for a fully complex-valued relaxation network, Neural Networks, 32, pp. 309-318, 2012.
  - R. Savitha, S. Suresh, and N. Sundararajan, Meta-cognitive learning in a fully complex-valued radial basis function network, Neural Computation, 24(5), pp. 1297-1328, 2012.

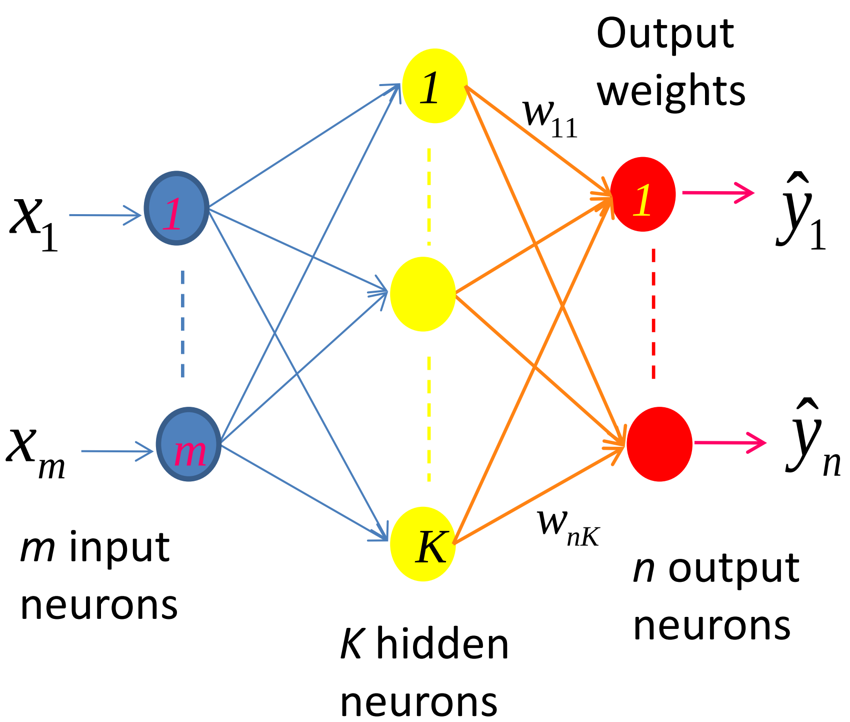
- G. Sateesh Babu, S. Suresh, Sequential projection based metacognitive learning in a Radial basis function network for classification problems, IEEE Trans, on Neural Networks and Learning Systems, 24(2), pp. 194-206, 2013.

META-COGNITIVE RBFN AND ITS SEQUENTIAL LEARNING ALGORITHM

McRBF: Schematic Diagram



McRBF: Cognitive Component



- Input Layer:
  - m neurons, linear
- Hidden Layer
  - K neurons, Gaussian

$$h_k^t = exp\left(-\frac{\|\mathbf{x}^t - \mu_k^l\|^2}{(\sigma_k^l)^2}\right)$$

- Output Layer
  - n neurons, linear

$$\hat{y}_j^t = \sum_{k=1}^{K} w_{kj} h_k^t$$

McRBF: Meta-cognitive component

- Monitory Signals
  - Predicted Class Label:

$$\widehat{c}^t = arg \max_{j \in 1, \cdots, n} \widehat{y}_j^t$$

  - Posterior Probability:

$$\hat{p}(j|\mathrm{x}^t) = \frac{min(1, max(-1, \hat{y}_j^t)) + 1}{2}, \quad j = c^t$$

  - Maximum Hinge Error:

$$E^t = \max_{j \in 1,2,\cdots,n} \left| e_j^t \right|$$

  - Class Specific Spherical Potential: $\psi \approx -\frac{2}{K} \sum_{k=1}^{K} h\left(\mathrm{x}^t, \boldsymbol{\mu}_k^l\right)$

McRBF: Meta-cognitive component

- Control signals
  - Sample Deletion Strategy
    - Remove similar samples as that of knowledge stored in the network
  - Sample Learning Strategy
    - Learn the current sample by any of the following way
      - Neuron Addition: Add new resource to capture novel knowledge
      - Neuron Deletion: Delete redundant resource
      - Parameter Update: Update existing knowledge
  - Sample Reserve Strategy
    - Current sample contain information but I will learn it later

Sequential learning algorithm

- Projection Based Learning for a Meta-cognitive Radial Basis Function Network (PBL-McRBFN):
  - What is Projection Based Learning?
    - Evolving learning algorithm
    - Classifier based on Hinge loss error function minimization
    - Based on the best human learning strategy, namely, self-regulated learning.
    - Uses past knowledge in learning
    - Fast learning algorithm:
      - Input parameters are initialized through meta-cognition
      - Output weights are estimated as a solution to a set of linear equations as a linear programming problem.

Projection Based Learning

- Hinge Loss Error Function

$$e_j^t = \begin{cases} 0 & \text{if } \; y_j^t \widehat{y}_j^t > 1 \\ y_j^t - \widehat{y}_j^t & \text{otherwise} \end{cases} \qquad j = 1, 2, \cdots, n$$

- Weight Minimization $J(W) = \dfrac{1}{2} \sum_{t=1}^{N} \sum_{j=1}^{n} (e_j^t)^2$

- Find optimal W $W^* = \arg \min_{W \in \Re^{K \times n}} J(W)$

Projection Based Learning-contd

- Minimum energy point is obtained using

$$\frac{\partial J(\mathbf{W})}{\partial w_{pj}} = 0, \quad p = 1, \cdots, K; \; j = 1, \cdots, n$$

- Solving the above equation, we have:

$$\sum_{k=1}^{K} \sum_{i=1}^{t} h_k^i h_p^i w_{kj} = \sum_{i=1}^{t} h_p^i y_j^i$$

- Which can be written as:

$$\sum_{k=1}^{K} a_{kp} w_{kj} = b_{pj}, \; p = 1, \cdots, K; \; j = 1, \cdots, n$$

- In matrix form, $\mathbf{AW} = \mathbf{B}$

- It can be shown that A inverse exists and hence

$$\mathbf{W}^* = \mathbf{A}^{-1}\mathbf{B}$$

Summary: PBL-McRBFN

- Initialization: Set sample 1 (t=1) as the first neuron (K=1)
- For samples t = 2, …, N
  - **Cognition:** Compute the output based on the current network
  - **Meta-cognition:**
    - **Monitoring**: Compute *significance of the sample* using maximum hinge-loss error, predicted class label, confidence of classifier, and class-specific spherical potential
    - **Control**: Choose suitable learning strategies
      - » Sample Deletion: Delete samples with insignificant knowledge
      - » Sample Learning
        - Neuron addition: Add a neuron (K = K+1) if the sample is very significant. Neuron addition threshold is self-regulated.
        - Parameter update: Update the output weight if the sample is significant. Parameter update threshold is self-regulated
      - » Sample reserve: Reserve the sample. Due to the self-regulatory nature of the thresholds, the sample may be used later.

Performance Evaluation

- Benchmark problems
  - Classification problems from UCI machine learning repository
- Applications
  - Whole brain image based Alzheimer's disease detection

Benchmark Problems: UCI machine learning repository

Datasets from UCI [8]

Datasets	#Features	#Classes	# Samples		Impact Factor	
			Train	Test	Train	Test
Image Segmentation (IS)	19	7	210	2100	0	0
Liver Disorder (LD)	6	2	200	145	0.17	0.14
Ionosphere (Ion)	34	2	100	251	0.28	0.28

For other benchmark problems and results, please refer to:
G Sateesh Babu, S Suresh, "Meta-cognitive RBF Network and its Projection Based Learning algorithm for classification problems," *Applied Soft Computing*, vol. 13, no. 1, pp. 654–666, 2013.

[8] Blake, C., & Merz, C. (1998). UCI repository of machine learning databases. Irvine: Department of Information and Computer Sciences, University of California, Irvine. http://archive.ics.uci.edu/ml/.

Results on Benchmark Problems

Data	PBL-McRBFN				SRAN				SVM		
	K	N_u	Testing		K	N_u	Testing		K	Testing	
			η_o	η_a			η_o	η_a		η_o	η_a
IS	50	89	94.2	94.2	47	113	92.3	92.3	127	91.4	91.4
LD	87	116	73.1	72.6	91	151	66.9	65.8	141	71	70.2
ION	18	58	96.4	96.5	21	86	90.8	91.9	43	91.2	88.5

For results to other benchmark problems and statistical studies, please refer to:

G Sateesh Babu, S Suresh, "Meta-cognitive RBF Network and its Projection Based Learning algorithm for classification problems," *Applied Soft Computing*, vol. 13, no. 1, pp. 654–666, 2013.

10 fold Cross validation results

Data set	Classifier	# Neurons		# Samples used		Testing					
						η_o		η_g		$F-measure$	
		Mean	Dev	Mean	Dev	Mean	Dev	Mean	Dev	Mean	Dev
HEART	SVM	44^a	7.39	70	0	77.3	2.69	77.35	2.56	0.76	0.02
	ELM	46.5	2.41	70	0	73.2	2.79	73.18	2.93	0.71	0.03
	PBL-McRBFN	28.2	2.39	56.9	8.54	**81.7**	1.13	**81.37**	1.02	0.79	0.01
LD	SVM	157.5^a	4.72	200	0	69.21	2.1	62.91	4.17	0.75	0.02
	ELM	127	15.67	200	0	64.55	3.8	63.52	4.04	0.68	0.04
	PBL-McRBFN	78.1	8.55	130.2	16.44	**70.82**	1.08	**69.87**	1.57	0.74	0.01
PIMA	SVM	252.7^a	42.28	400	0	76.76	1.45	66.78	6.58	0.57	0.07
	ELM	172	25.73	400	0	70.86	1.44	65.50	2.67	0.53	0.03
	PBL-McRBFN	100.8	8.67	185.9	29.58	**74.45**	2.12	**74.41**	1.00	0.63	0.01
BC	SVM	27.7^a	3.6	300	0	96.55	0.57	96.41	0.67	0.94	0.01
	ELM	37.9	1.34	300	0	96.97	0.47	96.89	0.87	0.95	0.01
	PBL-McRBFN	12.3	4.02	107.4	34.83	**97.77**	0.60	**97.95**	0.37	0.96	0.01
ION	SVM	70.9^a	10.27	100	0	91.24	1.11	90.10	2.69	0.87	0.02
	ELM	46	1.76	100	0	80.26	2.3	74.63	2.59	0.69	0.03
	PBL-McRBFN	20.6	3.68	47.5	11.14	**93.90**	1.31	**93.74**	1.78	0.91	0.01

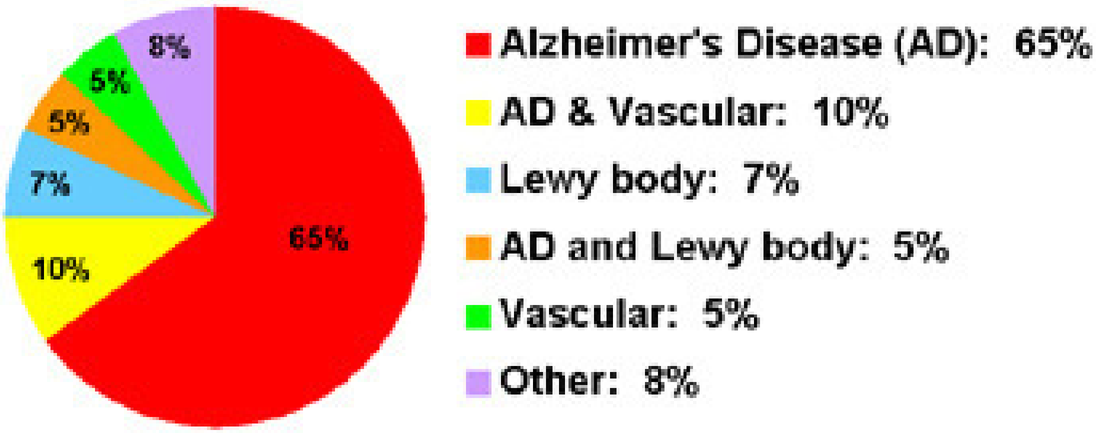
Observations

- First time in literature, human metacognition principles are integrated in machine learning framework.
- Self-regulation of cognitive component (RBF network) helps in achieving better generalization.
- Sample reserve strategy
  - Play vital role in Judgment of Learning
  - Use self-selected samples for validation of addition of neurons
  - One can also use it for preventing drift in sequential learning

Application – Whole Brain MR Imaging based Alzheimer's Disease Detection

Alzheimer’s disease

- Main threat to public health
- ~ 30 million AD patients worldwide
  - 3.7 million Indians
  - 5.3 million Americans
- AD - progressive, degenerative disease that leads to
  - memory loss, poor judgment and problems in learning
- At present, researchers know of no single cause nor of a cure



Diagnosis of Alzheimer's Disease

- **Clinical diagnosis criteria**
  - National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer's Disease and Related Disorder Association Criteria (NINDS-ADRDA) (http://m.medicalcriteria.com/crit/neuro_alzheimer.html)
  - Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV) (http://www.dnalc.org/view/2221-DSM-IV-criteria-for-Alzheimer-s-disease.html)

- **Neuropsychological testing**
  - Mini-Mental Status Examination (MMSE) (http://www.health.gov.bc.ca/pharmacare/adti/clinician/pdf/ADTI%20SMMSE-GDS%20Reference%20Card.pdf)
  - Clinical Dementia Rating (CDR) (http://www.biostat.wustl.edu/adrc/)

Neuro imaging in AD

- MRI - Magnetic Resonance Imaging
  - High spatial resolution
  - Exceptional soft tissue contrast
  - Can detect minute abnormalities
  - Can visualise and measure atrophy rates
- Advanced MR techniques
  - Diffusion Tensor Imaging - Tissue microstructure
  - Magnetic resonance spectroscopy - Brain metabolism
  - Functional MRI - Neural activity
- Early detection of AD from MRI is a promising alternative

Data Sets

- Open Access Series of Imaging Studies (OASIS) database (OASIS) (http://www.oasis-brains.org/)
  - 100 AD patients, 98 controls
  - Homogenous

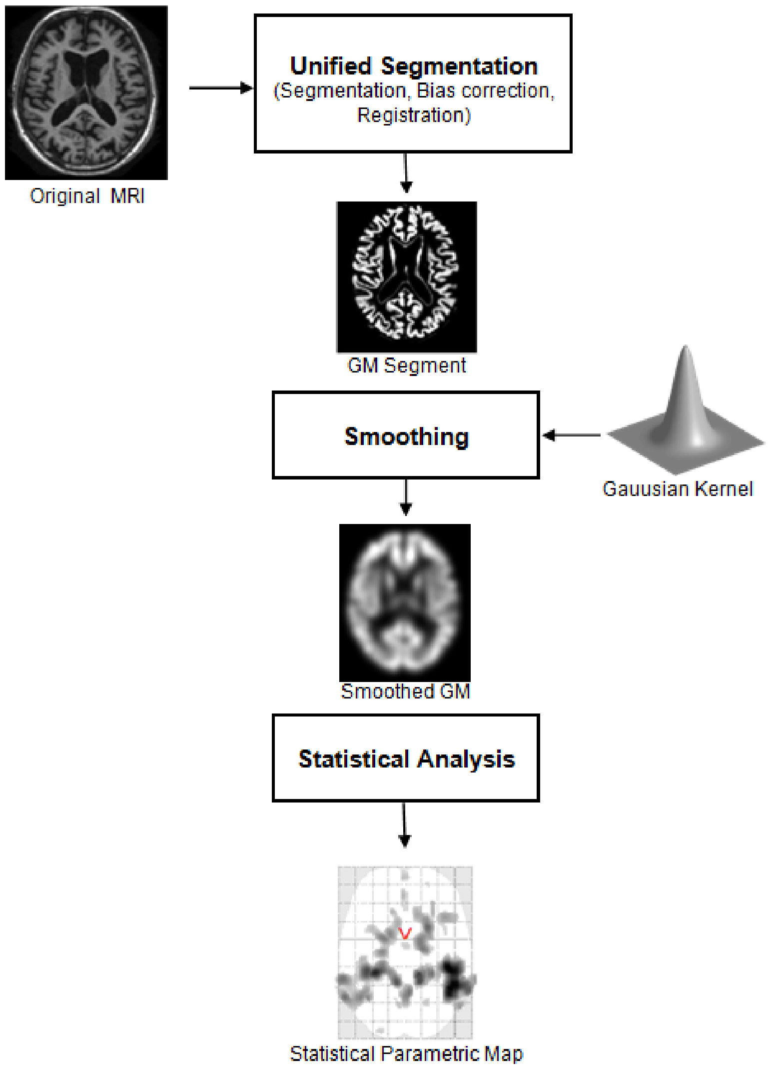
  Y. Fan et. al, "Integrated feature extraction and selection for neuroimage classification," *Medical Imaging 2009: Image Processing*, vol. 7259, no. 1, p. 72591U, 2009.

  W. Yang et. al, "ICA-based feature extraction and automatic classification of AD-related MRI data, *ICNC*, vol. 3, 2010, pp. 1261-1265.

- Alzheimer's Disease Neuroimaging Initiative (ADNI) (http://adni.loni.ucla.edu/)
  - 232 AD patients, 200 controls (as of Feb 2012)
  - Heterogenous

  W. Yang et. al, "ICA-based automatic classification of magnetic resonance images from ADNI data," *Part III, LSMS/ICSEE 10*, pp. 340-347, 2010.

Feature Extraction using VBM



- Voxel-based Morphometry (VBM) – image analysis technique

  (J. Ashburner and K. Friston, "Unified segmentation," NeuroImage, vol. 26, pp. 839–851, 2005.)

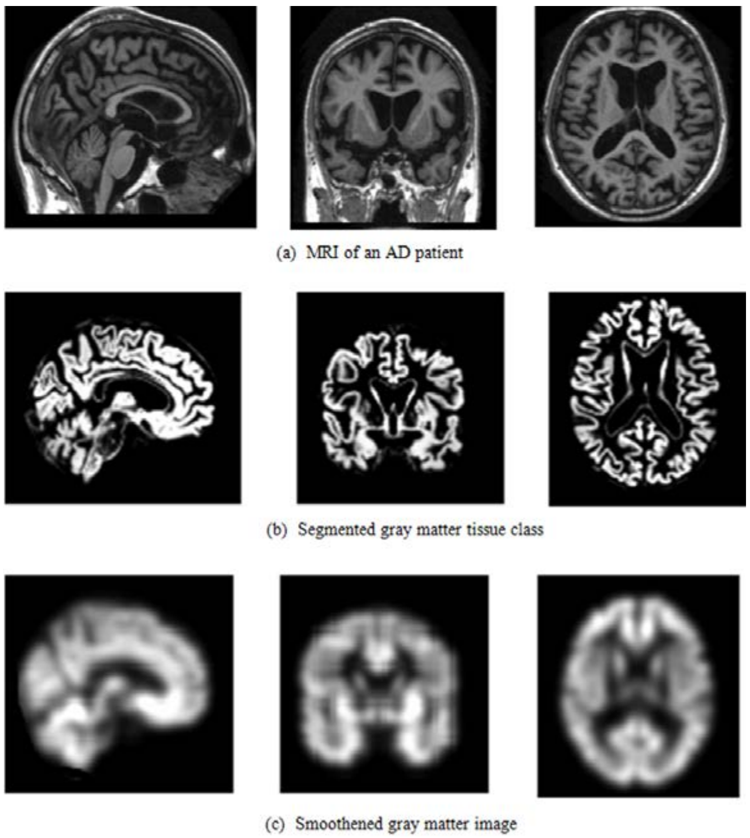
  – Identifies regional differences in gray or white matter between groups of subjects

  – Whole-brain analysis - does not require a priori assumptions about region of interest

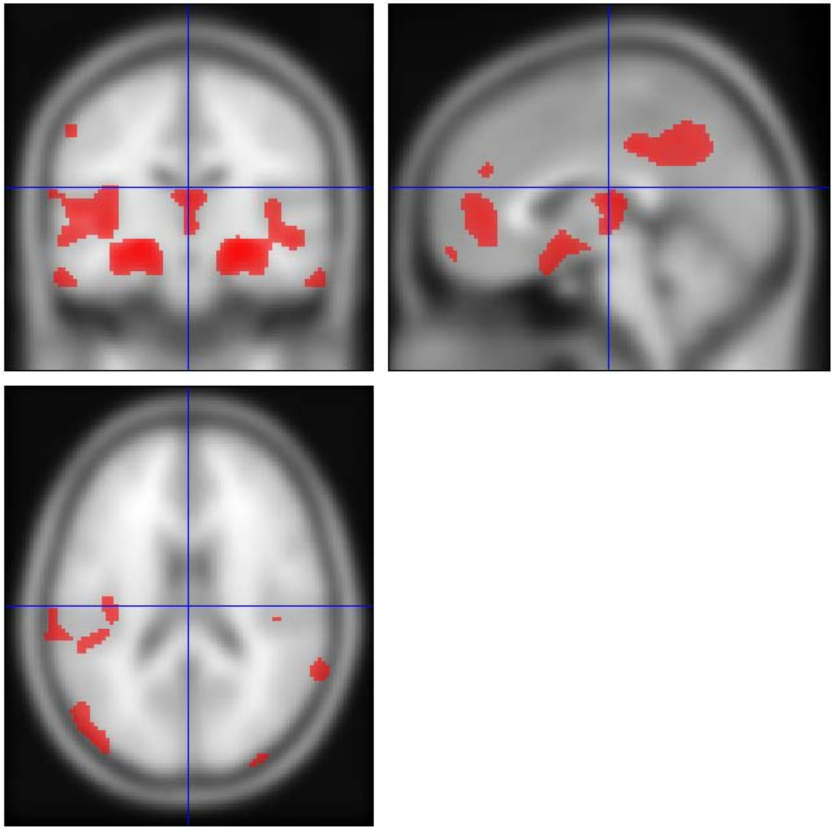
  – Fast and fully automated

MRI image and VBM results

Segmented and Smoothed images

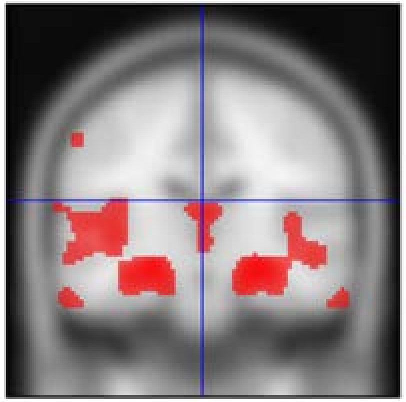


MIP from VBM

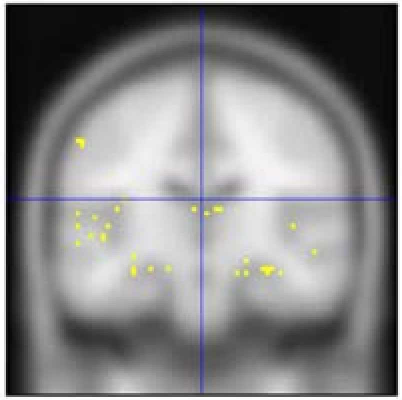


Imaging Biomarkers

Selected region



(a) VBM Detected 19879 voxels



(b) RFE Selected 906 voxels

Brain region names

Feature Type	No. of Features	Identified Regions
VBM	19879	Parahippocampal gyrus, Amygdala, Hippocampus, Superior temporal gyrus, Insula, Sub-gyrus, Precentral gyrus, Extra-nuclear
VBM + RFE	906	Parahippocampal gyrus, Hippocampus, Superior temporal gyrus, Insula, Precentral gyrus, Extra-nuclear

observations

- Results using PBL-McRBF RFE:
- Gender-wise analysis
  - Male – INSULA
    - Emotion and consciousness related problems
  - Female – PARAHIPPOCAMPAL GYRUS and EXTRA-NUCLUS regions
    - Memory encoding and retrieval related problems

- Age-wise analysis
  - 60-70 – SUPERIOR TEMPORAL GYRUS
    - Auditory related problems
  - 70-80 - PARAHIPPOCAMPAL GYRUS and EXTRA-NUCLUS regions
    - Memory encoding and retrieval related problems
  - 80 and above – HIPPOCAMPUS, LATERAL VENTICAL, PARAHIPPOCAMPAL GYRUS
    - Long-term memory, spatial navigation, memory coding and retrival

Conclusions

- For the first time, human metacognition principles are integrated in a machine learning framework.
- Self-regulation of cognitive component (RBF network) helps in achieving better generalization.
- McRBF effectively answer what-to-learn, when-to-learn and how-to-learn by
  - Sample deletion strategy
  - Sample learning strategy
    - addition/deletion and update
  - Sample reserve strategy

Other Applications using meta-cognitive neural networks

- **Medical Informatics:**
  - Alzheimer’s disease detection (Sateesh Babu et. al., ICML 2012/IEEE TNNLS 2013).
  - Parkinson’s disease detection (Sateesh Babu et. al., Expert Systems with Applications, 2013).

- **Video Analytics:**
  - Human Action Recognition (K. Subramanian et. al., International Journal of Neural Systems, 2012).
  - Human Emotion Recognition (K. Subramanian et. al., Submitted to IEEE TNNLS, 2013).

- **Complex-valued Signal Processing:**
  - Human Action Recognition using complex-valued features (R. V. Babu et. al., Neurocomputing, 2012)
  - QAM Equalization (R. Savitha et. al., Neural computation, 2012/IEEE TNNLS 2013)
  - Adaptive Beamforming (R. Savitha et. al., Neural computation, 2012/IEEE TNNLS 2013)

Part II
How metacognition is incorporated in PSO?

What is Optimization?

- The process of finding the conditions that give maximum or minimum of a function.
- The act of obtaining best results under given circumstances.
- The function to be optimized is called the "Objective Function" which consists of
  - Design Variables: Parameters for defining problem
  - Constraints: Bound the variables to certain values
- The Objective Function is a real function of *n* variables

$$f(x_1, x_2, \ldots, x_n)$$

Formulation of Optimization Problem

minimize the objective function

$$\min\ f(x), x = (x_1, x_2, \ldots\ldots, x_n)$$

subject to constraints

$$c_i(x) \geq 0$$

$$c_i(x) = 0$$

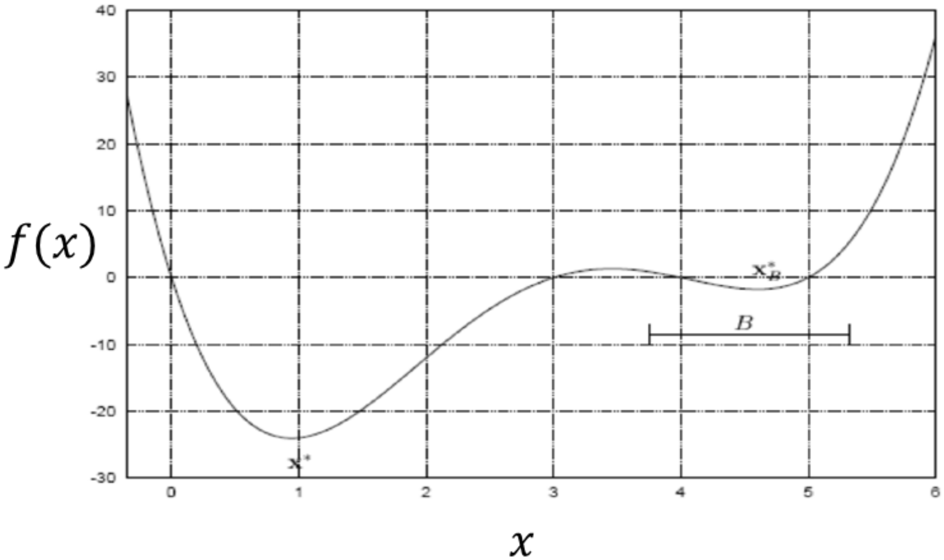
Example

$$\min\left[(x_1 - 2)^2 + (x_2 - 1)^2\right]$$

$$\text{subject}: x_1^{\ 2} - x_2^{\ 2} \leq 0$$

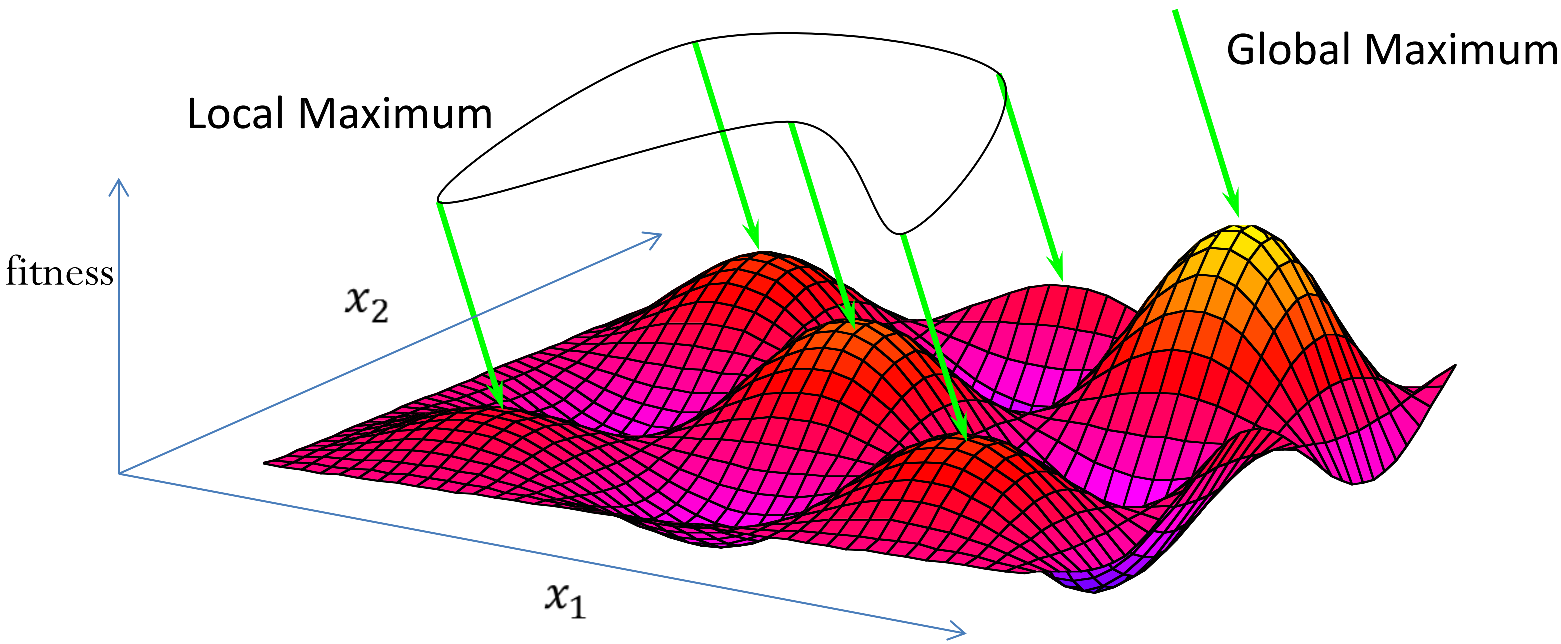
$$x_1 + x_2 \leq 2$$

Find the minimum of the function represented by



$$f(x) = x^4 - 12x^3 + 47x^2 - 60x$$

Global and local maximum



Constraint Black box optimization Problem where both x_1 and x_2 are bounded within [lb, ub].

Requirements of Optimization

- MODEL
  - Process of identifying Objective function, Variables and Constraints.
  - Mathematical representation of the problem
- ALGORITHM
  - Typically, complex models
  - Requires effective and reliable numerical algorithm
  - No universal optimization algorithm
  - Existing algorithms (not limited to)

1. Gradient Methods
2. Dynamic Programming
3. Non-Linear Programming
4. Evolutionary Algorithm
5. Quasi-Newton Method
6. ***Particle Swarm Optimization***

Particle Swarm Optimization

- PSO is a population based procedure to find the best possible solution.

(J. Kennedy and R.C. Eberhart. Particle swarm optimization. In Proceedings of IEEE Intl. Conf. on Neural Networks, pages 1942--1948, 1995.)

- Introduced in 1995 Dr. Eberhart and Dr. Kennedy.
- Motivated from the behavior of Bird Flock and Fish Schooling.
- The Scenario and Strategy used by birds can be summarized as:
  - Scenario
    - Birds searching for food
    - Searching for one piece
    - Only know how far food is
  - Strategy
    - Follow bird nearest to food

Particle Swarm Optimization

- In PSO each member is called as a 'Particle'
  - randomly initialized within the search space
  - Having fitness value (Objective Function)
  - And velocity (Directing Flight)
  - Flies around the search space.

- Flying is adjusted through
  - own flying experience (Exploration Process) and
  - flying experience of other particles (Exploitation Process).

Working of PSO Algorithm

PSO Update Equations

- The Velocity Update Equation:

$$V(t+1) = \omega . V(t) + c_1 r_1 (Pbest - X(t)) + c_2 r_2 (Gbest - X(t))$$

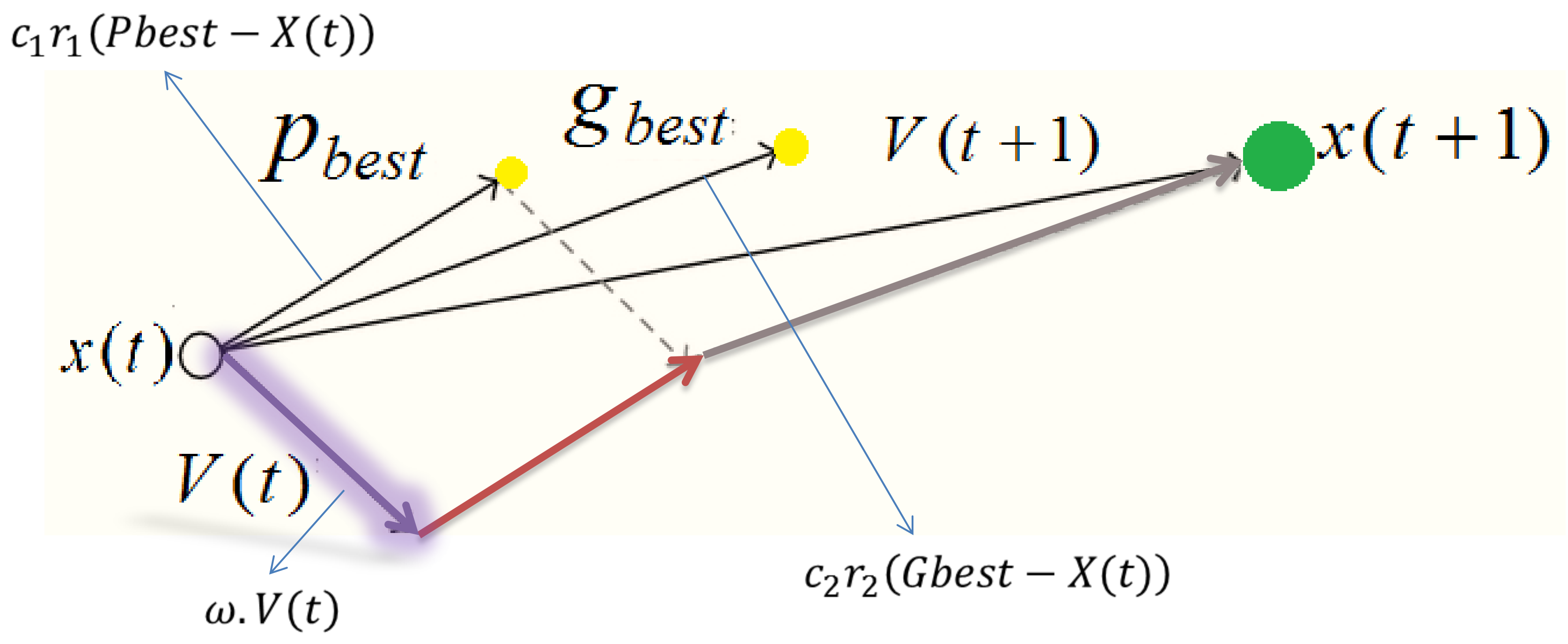
- The Position Update Equation:

$$X(t+1) = X(t) + V(t+1)$$

Where

- $V(t)$ & $X(t)$ are the current velocity and position
- ω is the inertia weight
- c_1 & c_2 are the acceleration constants
- r_1 & r_2 are random numbers
- $Pbest$ is the best position of each particle
- $Gbest$ is the global best position (Position of the best particle)

Update Mechanism



PSO Animation

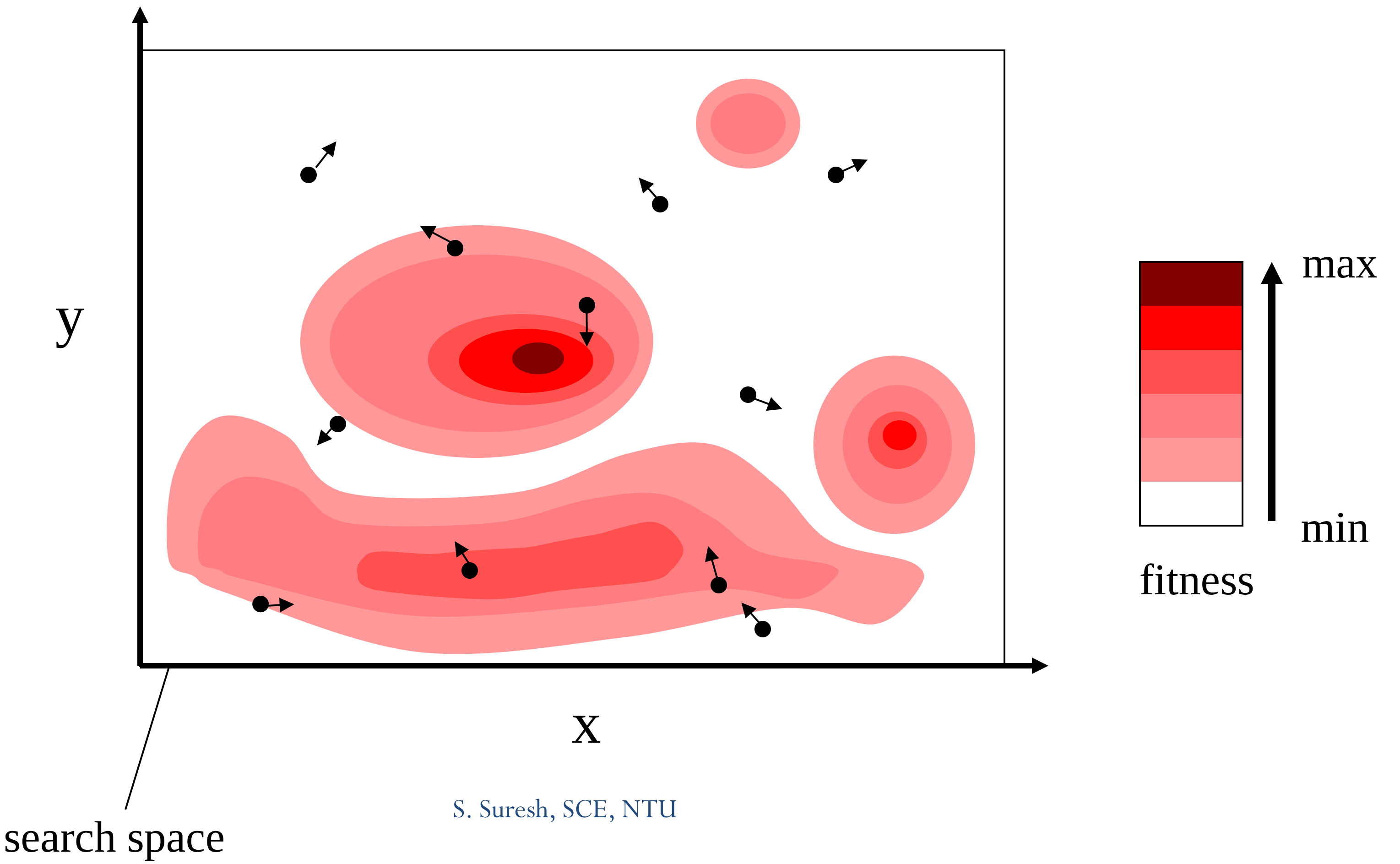
Performed by:

Research Group of Swarm Intelligence at Peking University

Ref: www.cil.pku.edu.cn/resources/**pso_and_itsvariants**

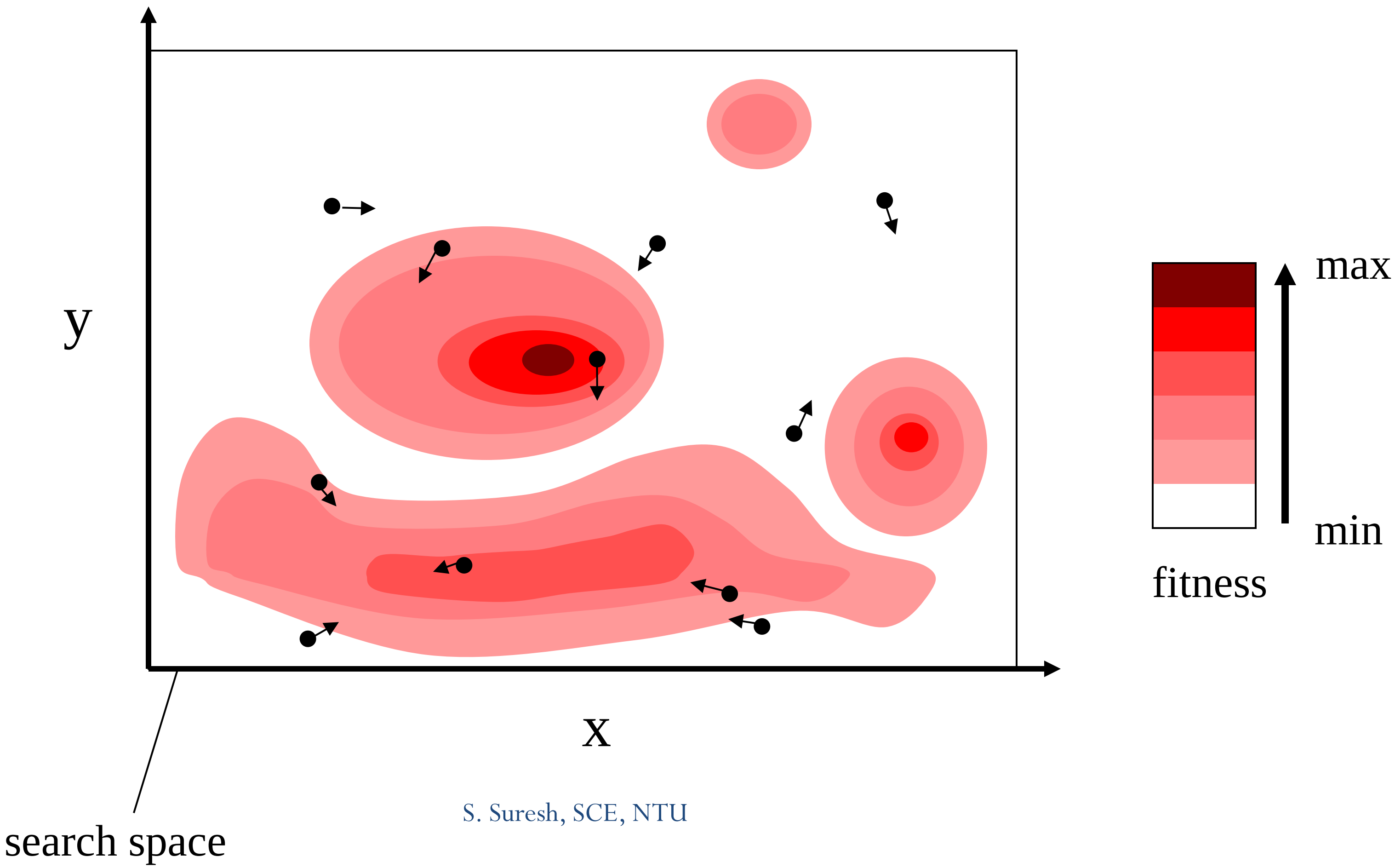
Initialization:

- Particles Randomly initialized in the search space
  - Position
  - Velocity



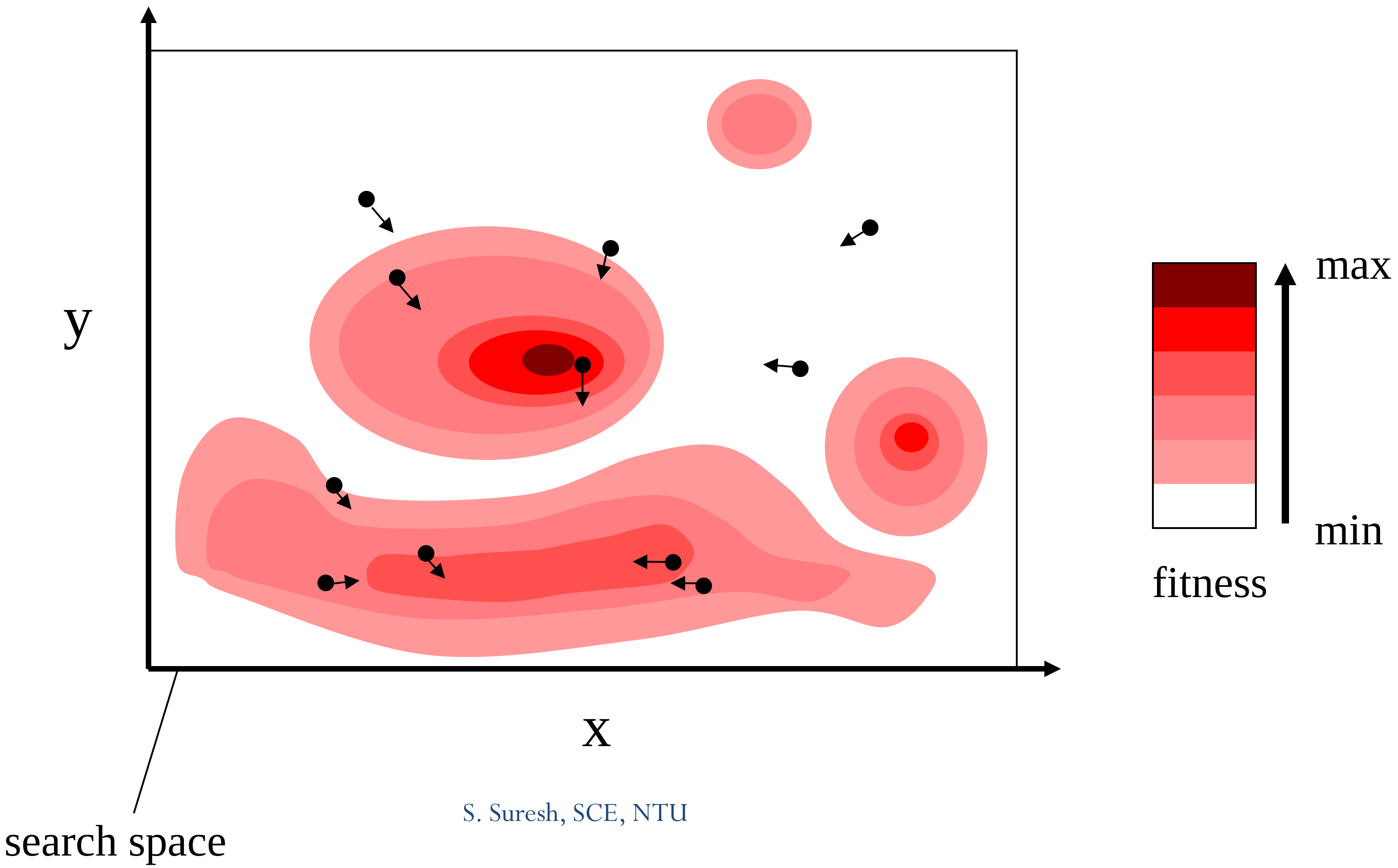
After Few Iteration:

- Particles started moving in the search space
  - Attracted to Local optima
  - Attracted to Global optima



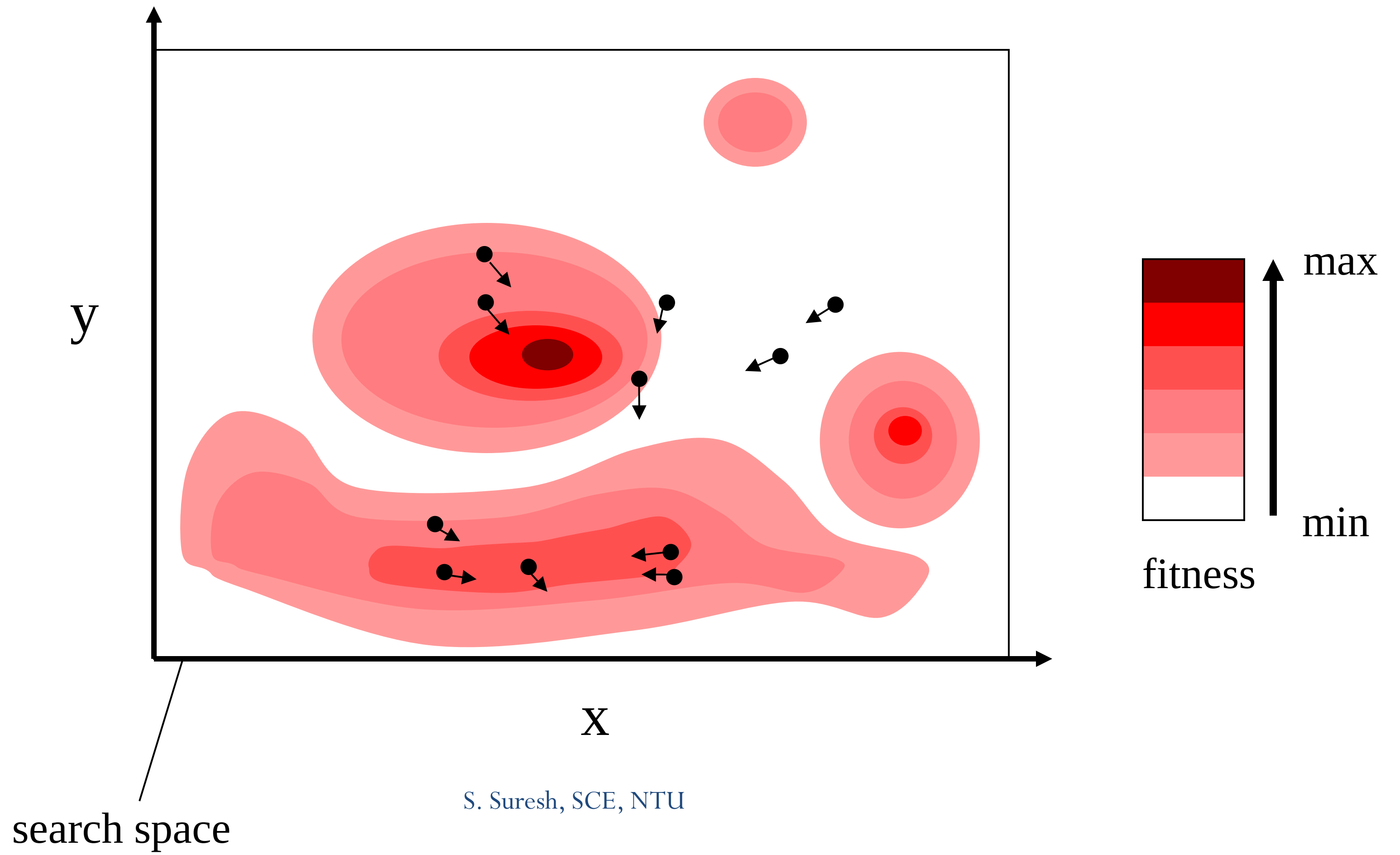
After Few Iteration:

- Particles moving in the search space
  - ➢ More are attracted to Global optima
  - ➢ Few are Converging to Local optima

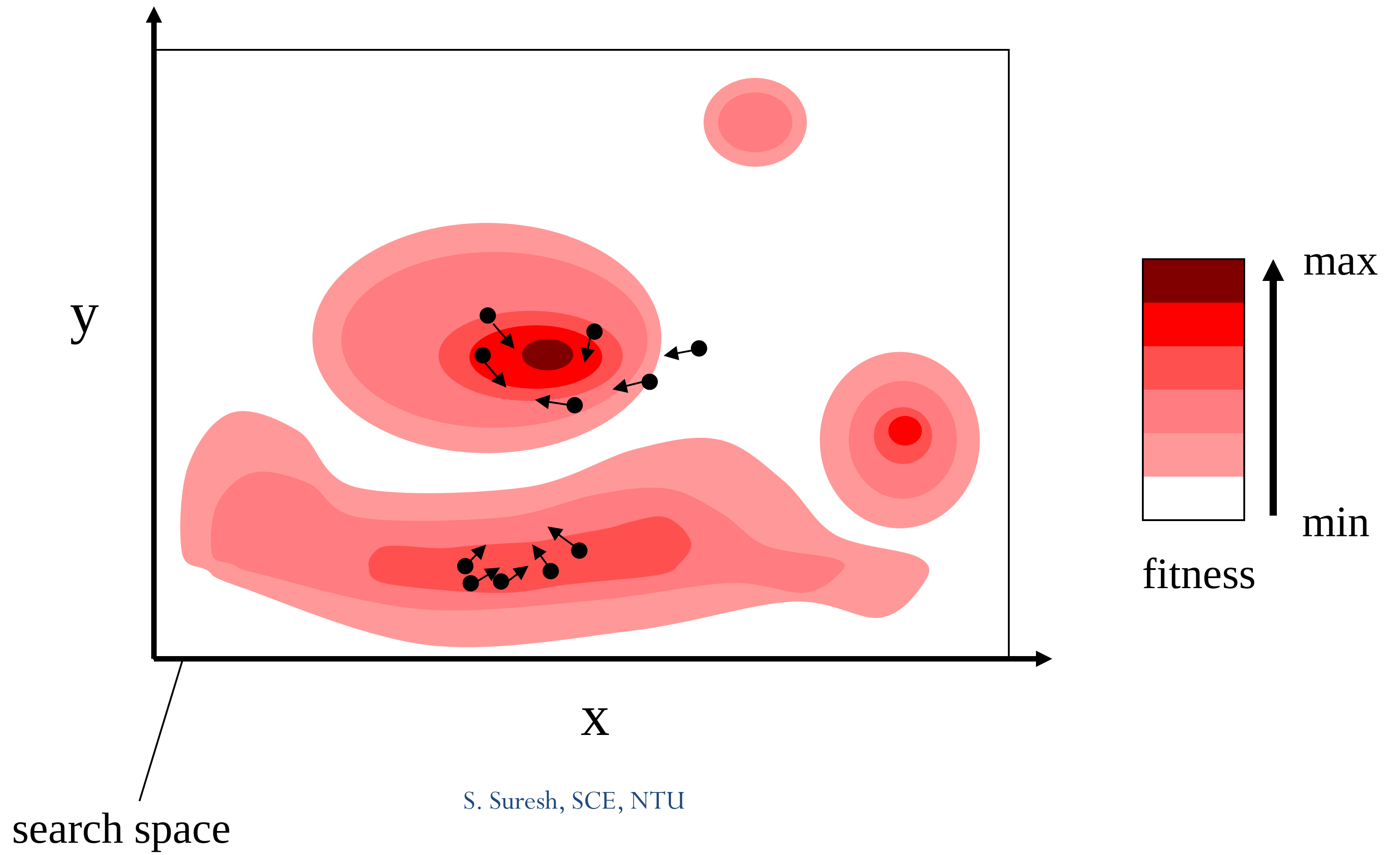


After Few more Iteration:

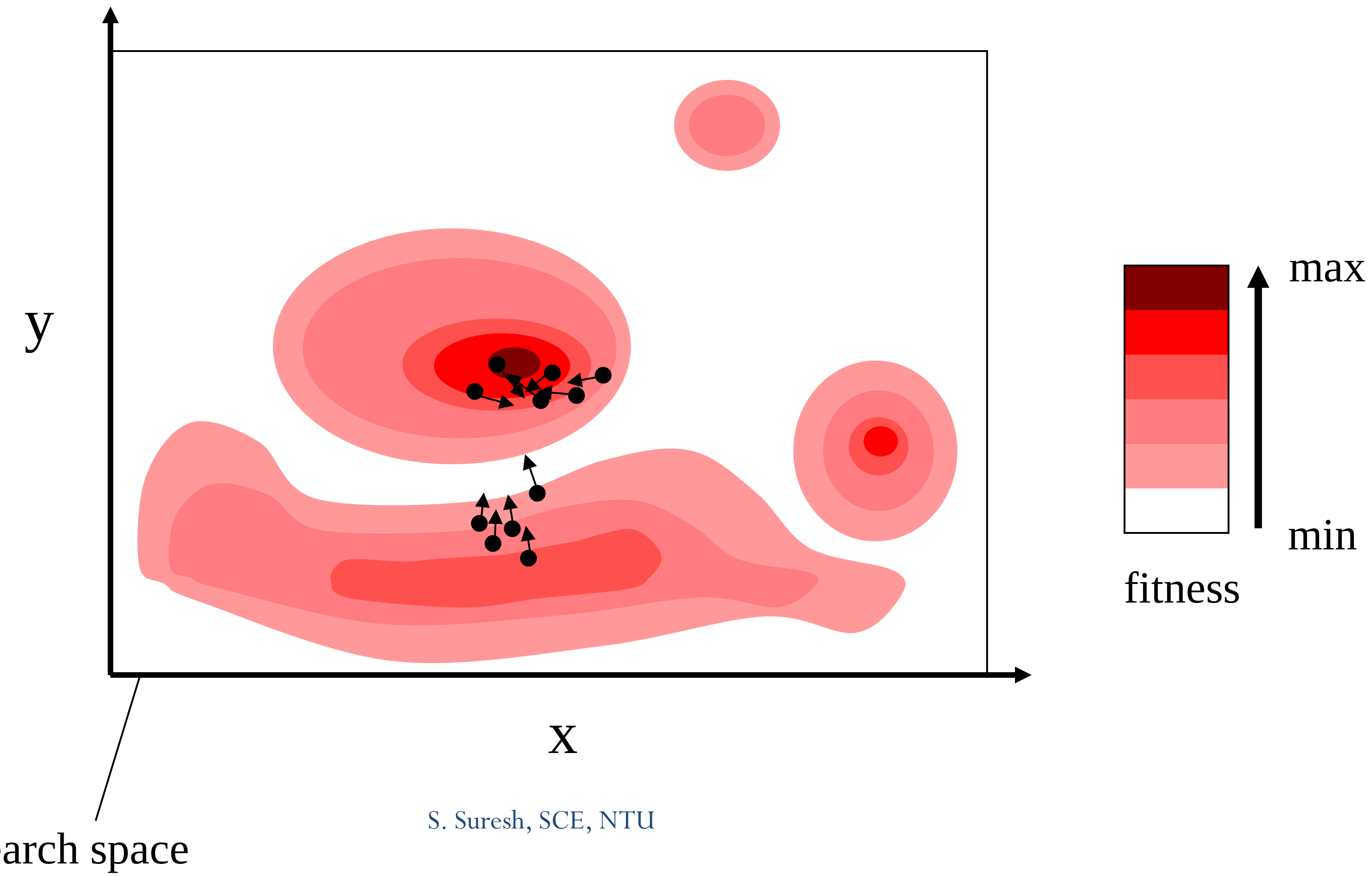
- Particles started moving in the search space
  - ➢ More particles converge to local optima
  - ➢ Particles are attracted to a Local optima



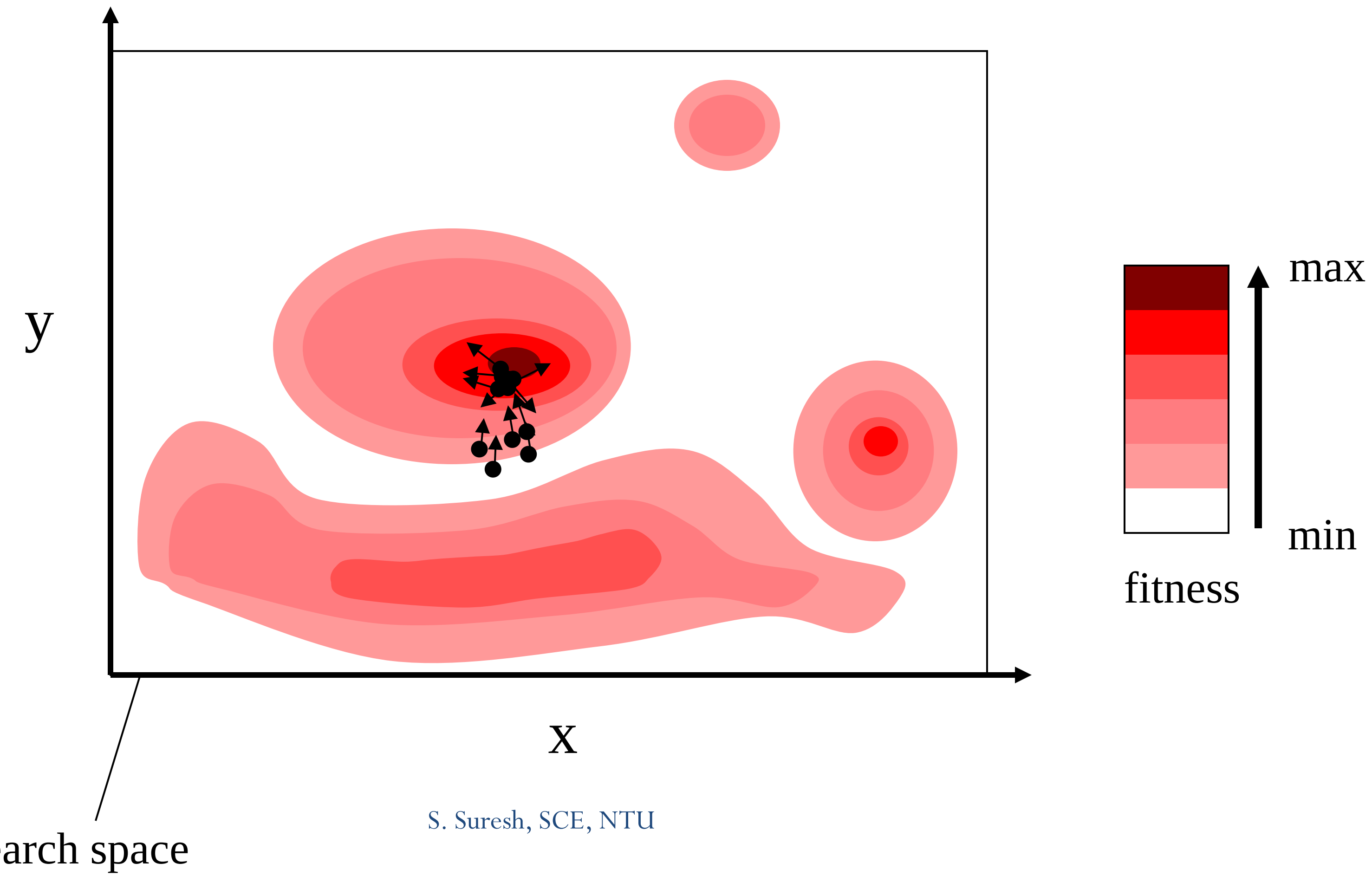
- Global Optima
  - Attracts the particles
  - Calls the particles from local optima



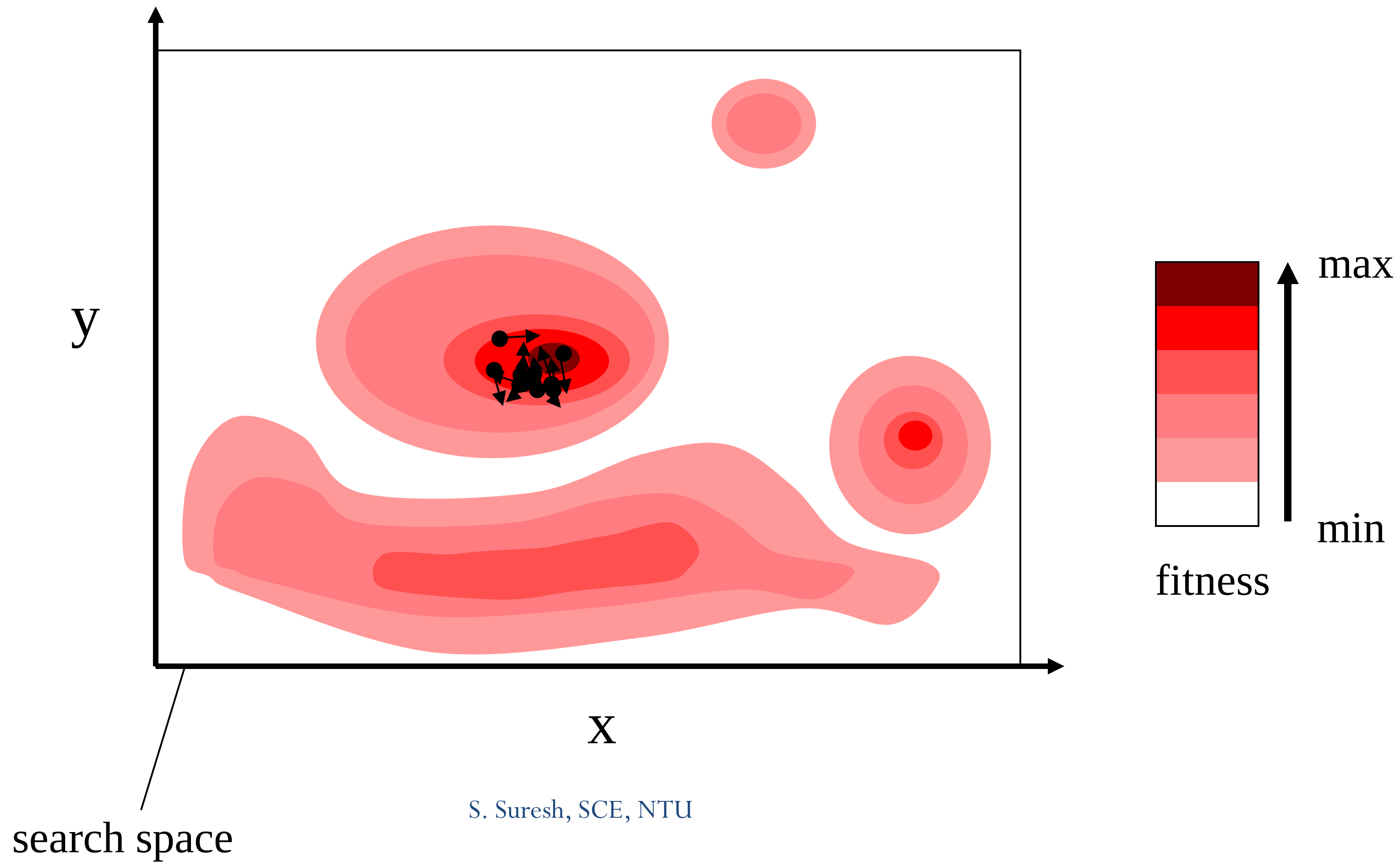
- Particles have started moving towards Global Optima



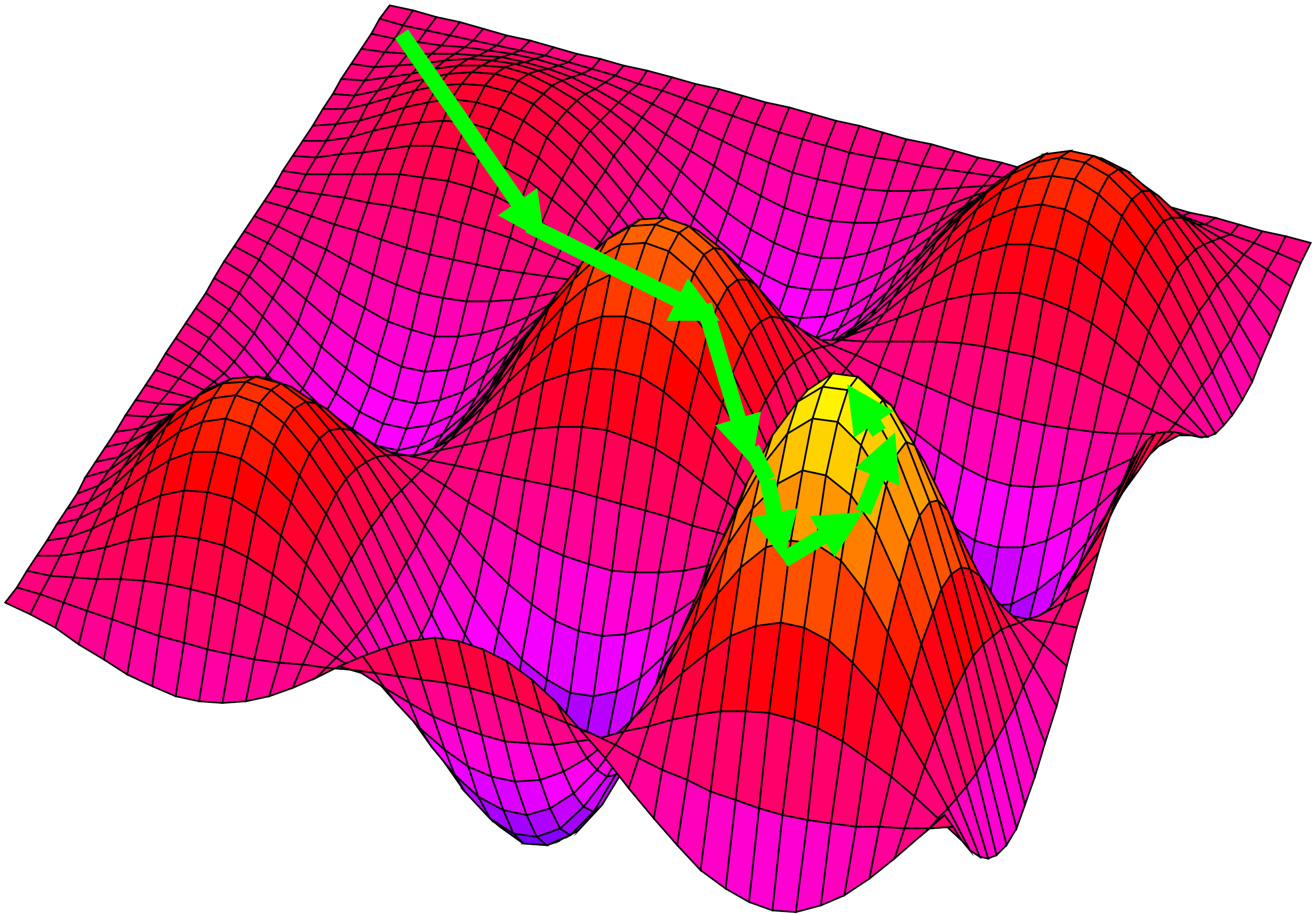
- Towards the end of Iterations
  - Particles are converging towards Global Optima



- Final Iteration
  - Finally the particles converge to a Global Optima



PSO Example



Path followed by a particle for convergence towards Global optimum

Variants of PSO Algorithm

- Discrete PSO ………………
  - can handle discrete binary variables
- MINLP PSO…………
  - can handle both discrete binary and continuous variables.
- Hybrid PSO………….
  - Utilizes basic mechanism of PSO and the natural selection mechanism, which is usually utilized by EC methods such as GAs.

Application

- PSO Can solve
  - Multi-objective optimization
  - Mixed integer programming
  - Difficult optimization problems
- Application
  - Complex structural design,
  - Aircraft wing design
  - Shape optimization
  - Finance application – Stock market
  - Power Transmission Network Expansion Planning (TNEP)

PSO-Summary

- Mathematical Simplicity
- Lesser Computational Efforts
- Premature Convergence
  – All the particles learn simultaneously from 'Pbest' and 'Gbest' even if they are far from the global optimum.
- *Main research areas in PSO*
  – *Parameter Setting*
  – *Deciding the neighborhood*
  – *Updating Learning Strategies*

NANYANG TECHNOLOGICAL UNIVERSITY

- Parameter Setting
  - Inertia weight (Shi & Eberhart, 1998)
  - Survey (Han et al., 2010)

$$V(t+1) = \omega.V(t) + c_1 r_1 (P_b - X(t)) + c_2 r_2 (G_b - X(t))$$

- Neighbourhood Topology
  - Dynamically changing (Liang et al., 2005)
  - Simultaneous search (Wang et al., 2011)

$$V(t+1) = \omega.V(t) + c_1 r_1 (P_b - X(t)) c_2 r_2 (\boxed{G_b} - X(t))$$

$\boxed{G_b \text{ is changed to } L_b}$

- Learning Strategy
  - Teaching and Peer learning (Lim & Isa, 2014)
  - Comprehensive Learning (Liang et al., 2006)

$$V(t+1) = \omega.V(t) + c_1 r_1 (P_b - X(t))$$

(CLPSO)

- Hybrid Version
  - Differential Evolution(Epitropakis et al., 2012)
  - Survey (Eslami et al., 2012)

All the research areas have provided better convergence characteristics to PSO.

Still exist the problem of Premature convergence.

Explore human learning principles for better performance improvement.

Self Regulating Particle Swarm Optimization Algorithm

Self Regulating Particle Swarm Optimization Algorithm (SRPSO)

- Research in human learning psychology (Nelson et al. 1990)
  - Humans are best planners
  - Self regulation leads towards better planning
  - Enables to decide
    - What to do, when to do and how to do.

- Best planner regulates his learning strategies
  - According to current state of knowledge and
  - Perception on the global knowledge

- The proposed algorithm:
  - Self Regulating Particle Swarm Optimization (SRPSO).

(M.R. Tanweer et. al., INS, 2015)

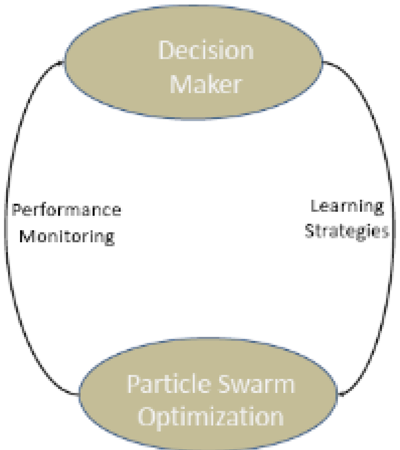


Fig: Effective decision making system for PSO analogous to Nelson and Naren's model (Nelson et al. 1990)

Strategies Introduced in SRPSO

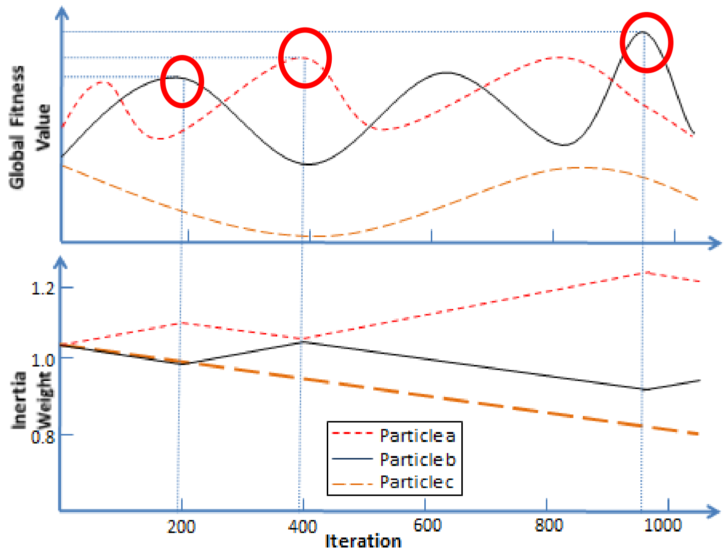
- Self-Regulated Inertia Weight
  - Used only for the best particle
  - Accelerates the exploration process
  - No self and social cognition
  - Inspired from the best learner: e.g. Hill Climber

- Self-Perception on Search Directions
  - Humans believe others based on trust
  - Trust defines the amount of information to be shared.
  - Partial social exploitation is used for all the other particles.

Strategies in SRPSO

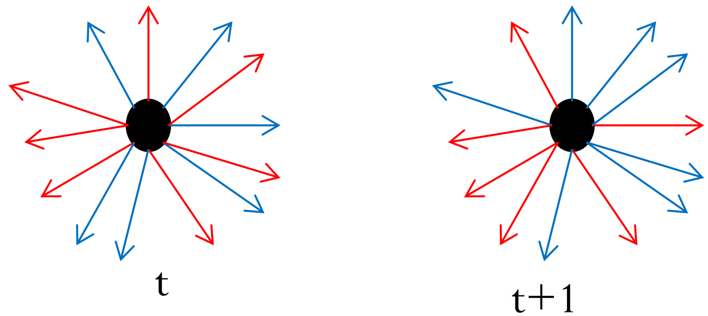
Self Regulating Inertia Weight:

- Best Particle
  - Increased Inertia weight
  - Enhanced exploration



Self Perception Strategy:

- Other Particles
  - Perception based selection of dimensions from global best directions
  - Exploration,
  - Self-exploitation and
  - Partial social-exploitation



Red arrows represent perception on search directions

SRPSO Update Equations

- The velocity update equation

$$V_{id}^{t+1} = \omega_i * V_{id}^{t} + c_1 r_1 P_{id}^{se}\left(P_{id}^{t} - X_{id}^{t}\right)$$
$$+ c_2 r_2 P_{id}^{so}\left(P_{gd}^{t} - X_{id}^{t}\right)$$

where

- ω_i is the self-regulated inertia weight
- P_{id}^{se} is the self-perception on own search directions
  - 0 for best particle and 1 for others
- P_{id}^{so} is the self-perception on global search directions

The position update equation is same as the basic PSO

Performance Evaluation

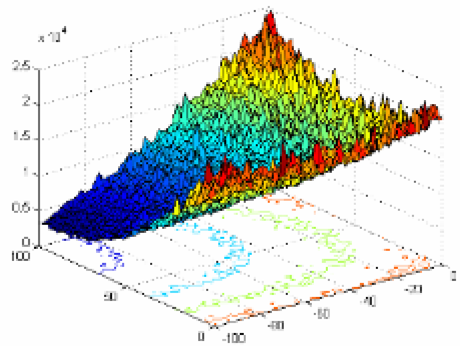
Experimental Setup

- CEC2005 Benchmark functions (Suganthan et al. 2005)
  - Unimodal (F1-F5)
  - Basic Multimodal (F6-F12)
  - Expanded Multimodal (F13-F14)
  - Hybrid Composition (F15-F25)
- Parameter settings
  - Inertia weight (Initial = 1.05, Final = 0.5)
  - Vmax = 0.07*Range
  - Swarm Size = 40
- Experiments
  - Functions evaluated for 100 times.
  - Mean, Median and Standard Deviation for 30 Dimension and 50 Dimension.

Analysis of Proposed strategies

Function:

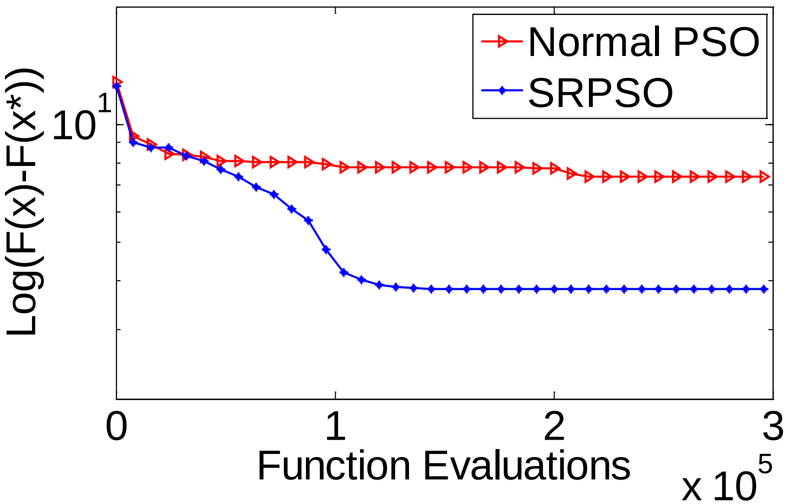
- F4 from CEC2005
  - Shifted Schwefel’s Problem with noise
  - Unimodal



Func.	Algorithm	Median	Mean	STD.
F4	Basic PSO	**1.044E+02**	**1.544E+02**	9.842E+01
	SRPSO-w (η=0.5)	9.495E+01	1.028E+02	5.577E+01
	SRPSO-w (η=1)	**8.079E+01**	**1.001E+02**	6.607E+01
	SRPSO-w (η=2)	8.874E+01	1.011E+02	6.046E+01
	SRPSO-p (λ=0.25)	1.339E+02	1.393E+02	6.071E+01
	SRPSO-p (λ=0.5)	5.685E+01	6.889E+01	4.094E+01
	SRPSO-p (λ=0.75)	**4.560E+01**	**6.191E+01**	4.891E+01
	SRPSO (η=1 & λ=0.5)	**1.998E+01**	**2.988E+01**	2.480E+01

Individual Strategy and Combined Effect:

- SRPSO-w: SRPSO only with self-regulated inertia weight
- SRPSO-p: SRPSO only with self-perception
- η is taken as ‘1’ because it provides best solution.
- λ is selected as 0.5 because
  - Lower value will make SRPSO like basic PSO.
  - Higher value will make particles fly in their own directions.



Convergence:

- SRPSO has significantly improved the performance of PSO.
- Similar observations are made for all the CEC2005 benchmark functions.

Selected PSO variants for Comparison

- Clerc et al. "The particle swarm - Explosion, stability, and convergence" (2002). (χPSO)
  - Introduced Constriction factor (χ)
  - Better convergence on selected problems (Bui et al., 2010)
- Mendes et al. "The fully informed particle swarm: simpler, maybe better" (2004). (FIPS)
  - New information flow scheme
  - Better convergence on multimodal functions (Chen et al., 2013)
- Liang et al. "Dynamic multi-swarm particle swarm optimizer" (2005). (DMS-PSO)
  - Dynamically changing neighborhood
  - Slower convergence speed (Nasir et al., 2012)
- Kennedy. "Bare bones particle swarms" (2003). (BBPSO)
  - Gaussian search space
  - Better performance on Hybrid Composition functions (Blackwell, 2012)
- Parsopoulos. "On the computation of all global minimizers through PSO" (2004). (UPSO)
  - Combined effect of global and local variants
  - Better convergence on selected problems (Epitropakis et al., 2012)
- Liang et al. "Comprehensive learning PSO for global optimization of multimodal" (2006). (CLPSO)
  - Particles learn from self-cognition
  - Better convergence on complex multimodal functions (Nasir et al., 2012)

Results (Unimodal)

Function	Algorithm	30 Dimensions			50 Dimensions		
		Median	Mean	STD.	Median	Mean	STD.
F1	χPSO	5.328E+00	9.657E+00	1.233E+01	5.328E+00	9.657E+00	1.233E+01
	BBPSO	**0.000E+00**	**0.000E+00**	0.000E+00	**0.000E+00**	**0.000E+00**	0.000E+00
	DMSPSO	1.143E+02	3.135E+02	4.149E+02	2.349E+02	3.870E+02	3.855E+02
	FIPS	3.185E+02	5.252E+02	5.571E+02	1.149E+03	1.673E+03	1.524E+03
	UPSO	1.269E+03	1.306E+03	7.328E+02	6.840E+02	7.100E+02	3.290E+02
	CLPSO	**0.000E+00**	**0.000E+00**	0.000E+00	**0.000E+00**	**0.000E+00**	0.000E+00
	SRPSO	**0.000E+00**	**0.000E+00**	0.000E+00	**0.000E+00**	**0.000E+00**	0.000E+00
F2	χPSO	**0.000E+00**	1.573E+01	8.112E+01	2.334E+02	7.774E+02	1.806E+03
	BBPSO	6.000E-03	9.260E-03	8.480E-03	2.407E+02	2.886E+02	1.452E+02
	DMSPSO	1.536E+02	7.801E+02	2.109E+03	3.311E+02	9.666E+02	1.409E+03
	FIPS	1.460E+04	1.470E+04	2.316E+03	2.633E+04	2.574E+04	4.424E+03
	UPSO	6.688E+03	7.602E+03	5.290E+03	3.632E+03	4.220E+03	2.894E+03
	CLPSO	3.828E+02	3.828E+02	1.060E+02	1.013E+04	1.021E+04	1.357E+03
	SRPSO	**0.000E+00**	**0.000E+00**	0.000E+00	**1.400E-03**	**3.767E-03**	1.753E-02
F3	χPSO	3.491E+06	1.020E+07	1.336E+07	1.852E+07	1.988E+07	1.266E+07
	BBPSO	1.243E+06	1.295E+06	5.728E+05	3.693E+06	3.709E+06	9.352E+05
	DMSPSO	3.898E+06	5.623E+06	6.225E+06	8.835E+06	1.317E+07	1.579E+07
	FIPS	1.530E+07	1.945E+07	1.109E+07	5.586E+07	5.867E+07	2.346E+07
	UPSO	4.308E+07	5.303E+07	3.856E+07	4.885E+07	5.340E+07	3.743E+07
	CLPSO	1.204E+07	1.188E+07	3.107E+06	5.084E+07	4.930E+07	1.161E+07
	SRPSO	**2.542E+01**	**4.519E+01**	5.820E+01	**6.967E+02**	**5.316E+03**	2.321E+04

3 Algorithms have converged to the optimum solution

SRPSO has converged to the optimum solution

SRPSO has provided much better results than the other algorithms

Results (Multimodal)

Func.	Algorithm	30 Dimensions			50 Dimensions		
		Median	Mean	STD.	Median	Mean	STD.
F6	χPSO	3.602E+02	1.170E+03	1.790E+03	3.979E+01	6.370E+06	2.129E+07
	BBPSO	1.148E+01	2.796E+01	4.218E+01	4.016E+01	5.831E+01	4.576E+01
	DMSPSO	2.226E+06	2.721E+07	7.289E+07	2.226E+06	1.768E+07	4.103E+07
	FIPS	9.832E+06	2.457E+07	3.493E+07	6.483E+07	8.021E+07	6.118E+07
	UPSO	6.826E+06	1.187E+07	1.355E+07	1.160E+06	2.731E+06	3.667E+06
	CLPSO	**7.369E+00**	**1.779E+01**	2.285E+01	8.998E+01	8.705E+01	3.757E+01
	SRPSO	1.472E+01	3.978E+01	5.708E+01	**3.053E+01**	**5.008E+01**	4.440E+01
F7	χPSO	6.788E+03	6.780E+03	1.291E+02	6.158E+03	6.154E+03	7.416E+01
	BBPSO	4.696E+03	4.696E+03	5.039E-01	6.195E+03	6.197E+03	4.553E+00
	DMSPSO	4.297E+03	4.335E+03	2.190E+02	6.029E+03	6.050E+03	1.312E+02
	FIPS	7.507E+03	7.477E+03	2.158E+02	1.037E+04	1.036E+04	2.122E+02
	UPSO	7.513E+03	7.524E+03	3.409E+02	7.419E+03	7.420E+03	3.034E+02
	CLPSO	4.696E+03	4.696E+03	1.837E-12	6.195E+03	6.195E+03	4.594E-12
	SRPSO	**2.940E-02**	**7.610E-02**	9.962E-02	**5.569E-01**	**5.444E-01**	2.253E-01
F12	χPSO	1.128E+04	1.872E+04	2.137E+04	2.896E+05	3.293E+05	1.922E+05
	BBPSO	**1.590E+03**	2.585E+03	2.746E+03	1.447E+04	1.505E+04	9.270E+03
	DMSPSO	5.375E+04	7.843E+04	6.836E+04	1.212E+05	1.659E+05	1.461E+05
	FIPS	4.679E+04	5.185E+04	3.213E+04	2.771E+05	2.929E+05	1.490E+05
	UPSO	7.752E+04	8.984E+04	5.430E+04	6.052E+04	7.135E+04	4.785E+04
	CLPSO	1.293E+04	1.324E+04	4.162E+03	8.996E+04	8.949E+04	2.001E+04
	SRPSO	1.642E+03	**2.495E+03**	2.804E+03	**6.996E+03**	**1.183E+04**	1.215E+04

Best Solution in 50 Dimension

Better Solutions several order better than all other variants

Better mean performance in both dimension.

Results (Hybrid Composition)

Func.	Algorithm	30 Dimensions			50 Dimensions		
		Median	Mean	STD.	Median	Mean	STD.
F16	χPSO	1.512E+02	1.867E+02	1.055E+02	1.679E+02	1.872E+02	5.268E+01
	BBPSO	1.187E+02	**1.356E+02**	5.300E+01	1.318E+02	**1.408E+02**	3.960E+01
	DMSPSO	2.250E+02	2.916E+02	1.683E+02	1.419E+02	1.670E+02	8.193E+01
	FIPS	3.271E+02	3.414E+02	1.081E+02	3.227E+02	3.296E+02	6.967E+01
	UPSO	3.827E+02	3.865E+02	1.416E+02	2.909E+02	3.287E+02	1.336E+02
	CLPSO	1.413E+02	1.453E+02	3.171E+01	1.967E+02	1.969E+02	3.751E+01
	SRPSO	**8.388E+01**	1.783E+02	1.723E+02	**8.372E+01**	1.504E+02	1.276E+02
F19	χPSO	9.761E+02	9.355E+02	6.813E+01	9.346E+02	9.383E+02	1.234E+01
	BBPSO	9.247E+02	9.209E+02	3.222E+01	9.866E+02	9.913E+02	1.812E+01
	DMSPSO	9.191E+02	9.328E+02	2.703E+01	9.297E+02	9.315E+02	9.374E+00
	FIPS	1.047E+03	1.049E+03	1.873E+01	1.070E+03	1.070E+03	1.603E+01
	UPSO	1.040E+03	1.049E+03	4.636E+01	1.027E+03	1.028E+03	3.344E+01
	CLPSO	9.140E+02	9.102E+02	1.853E+01	9.435E+02	9.418E+02	1.317E+01
	SRPSO	**8.281E+02**	**8.282E+02**	1.619E+00	**8.455E+02**	**8.461E+02**	3.846E+00
F25	χPSO	1.750E+03	1.750E+03	7.509E+00	1.682E+03	1.682E+03	5.316E+00
	BBPSO	1.669E+03	1.668E+03	7.609E+00	1.724E+03	1.724E+03	6.689E+00
	DMSPSO	1.639E+03	1.640E+03	8.368E+00	1.675E+03	1.676E+03	4.880E+00
	FIPS	1.780E+03	1.781E+03	1.046E+01	1.866E+03	1.866E+03	7.076E+00
	UPSO	1.778E+03	1.778E+03	1.256E+01	1.769E+03	1.771E+03	1.389E+01
	CLPSO	1.659E+03	1.659E+03	4.102E+00	1.701E+03	1.702E+03	2.610E+00
	SRPSO	**1.247E+03**	**1.113E+03**	3.227E+02	**1.288E+03**	**1.292E+03**	3.059E+01

Better median performance

Better performance in both dimension

Better mean performance in both dimension.

Results Analysis

- Better solutions in both 30D and 50D cases.
- Best solutions
  - All unimodal functions.
  - 4 out of 7 basic multimodal functions.
  - 1 of the 2 expanded multimodal functions.
  - 8 out of 11 Hybrid composition functions.

Summary

- PSO is a simple and effective optimization algorithm.
- It experiences premature convergence.
- It has been extensively researched.
- Incorporating human learning principles in PSO is a new search direction.
- SRPSO is human self-learning inspired PSO variant.
- SRPSO has significantly enhanced PSO convergence characteristics.
- Faster convergence closer to optimum solution has been observed.

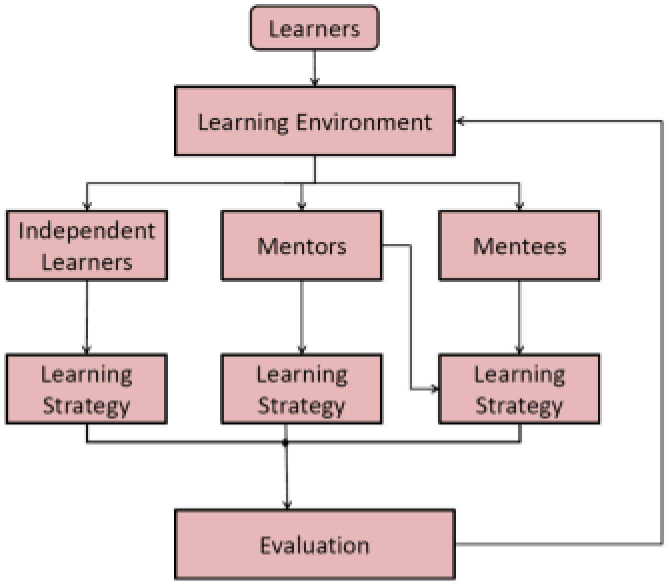
Limitations in SRPSO

- Only incorporates Human self-cognition
- Same perception for all particles
- Diversity management is not proper
- Performance suffers on few functions
- **Need of addressing Human social behaviour**

Mentoring based Particle Swarm Optimization Algorithm

Motivation

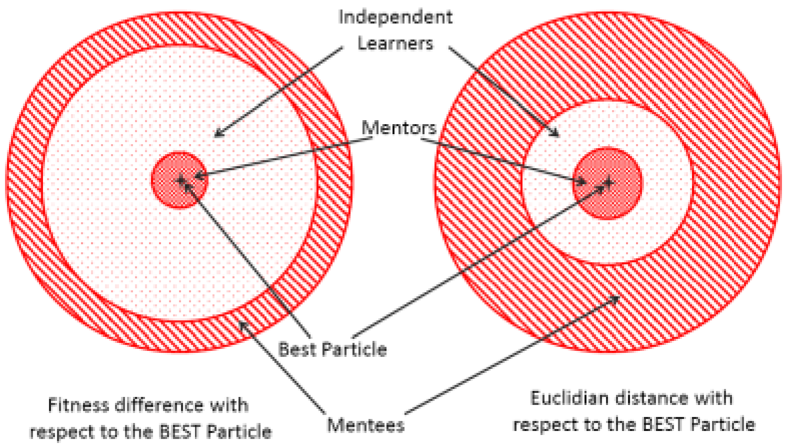
- Human social learning in SLPSO (Cheng & Jin, 2015)
- Human self-cognition in SRPSO (Tanweer et al., 2015)
- Need to address both self and social learning together
- Mentoring based learning
  - Process of positive learning w
  - Dynamic Learning environme
  - Effective Learners act as Men for less efficient learners, the
  - Moderate Learners perform independent learning
  - Performance decides the lear



Mentoring based Particle Swarm Optimization (MePSO) Algorithm

- Human Mentoring process incorporated in PSO.
- Particles are divided into three groups
  - Mentors: Top performing particles
  - Mentees: Least performing particles
  - Independent Learners: All other particles

$$i = \begin{cases} M_r, & if\ S_f \leq 5\%\ and\ S_{ed} \leq 10\% \\ M_e, & if\ S_f > 90\%\ and\ S_{ed} > 50\% \\ I_n\ , & if\ 5\% < S_f \leq 90\%\ and\ 10\% < S_{ed} \leq 50\% \end{cases}$$

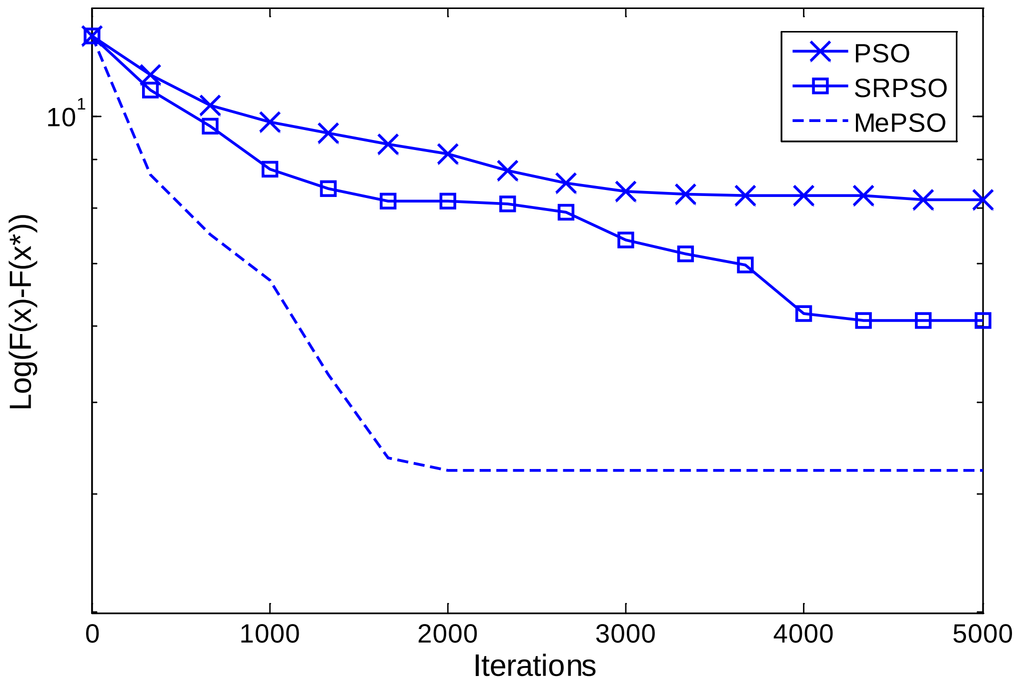


Learning Strategies

- Mentors Group
  - Higher self-cognition and partial social cognition
  - Best Particle: Self-regulating inertia weight from SRPSO (Tanweer et al., 2015)
- Mentee Group
  - Either mentor or self guidance
- Independent Learners Group
  - Self-perception strategy (Tanweer et al., 2015)

Convergence Analysis: Impact of Mentoring

- A unimodal Rotated Discus Problem (F_4) (Liang et al., 2013)
  - Performance comparison among PSO, SRPSO and MePSO



Faster Convergence

Closer to true optima

Real World Application

Problem Definition

Transmission Network Expansion Planning (TNEP) problem

(I de J Silva et. al., IET Proceedings- Generation Transmission and Distribution, Dec 2005)

- Determine the set of new lines to be constructed
  - Cost of Expansion is minimum
  - No overload
- The Problem set includes:
  - Generating Points, generating capacity and voltage level
  - Load point and load value
  - Existing lines and transformer units
  - Investment cost of lines, power rating and transformer
  - Power losses cost
- The dynamic formulation is a large scale non-linear mixed interger optimization problem.

Problem Formulation

TNEP without security constraints

(I de J Silva et. al., IET Proceedings, Dec 2005)---[a]

(Das & Suganthan, Tech. Report CEC2011, Dec 2010)---[b]

$$\min v = \sum_{l \in \Omega} c_l n_l$$

Where

- c_l is the cost of line added in l^{th}right-of-way
- n_l is the number of circuits added in l^{th}right-of-way

(Further details about the problem are available in [a] and [b])

Problem Formulation

The cost function for each solution is

$$f = \sum_{l \in \Omega} c_l n_l + W_1 \sum_{ol} \left(abs(f_l) - \overline{f_l} \right) + W_2 (n_l - \overline{n_l})$$

where

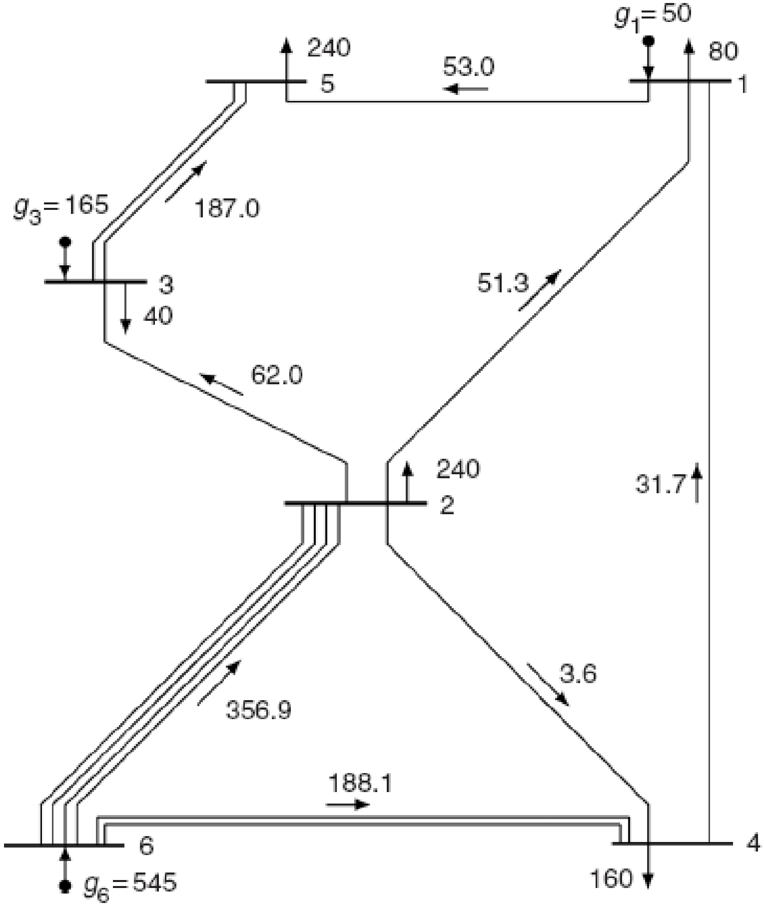
- 'ol' are the set of overload lines
- f_l is the total real power flow by the circuit in l^{th} right-of-way
- $\overline{f_l}$ is the maximum allowed real power flow in the circuit in l^{th} right-of-way
- $\overline{n_l}$ is the maximum number of circuits that can be added in l^{th} right-of-way
- W_1 and W_2 are the constants.

In the cost function

- 1st term is the total investment cost
- 2nd and 3rd terms are added to handle the violation of power flow constraints and maximum number of circuits respectively.

EXAMPLE

- **GARVER SYSTEM:** The Garver system has six buses, 15 candidate branches, a total demand of 760MW, and a maximum possible number of added lines per branch equal to five.
  - Solved using Chu-Beasley GA
  - Optimal Cost = USD 200 000
  - Following lines are added
    - $n_{2-6} = 4$
    - $n_{3-5} = 1$
    - $n_{4-6} = 2$



Garver system optimal plan without generation rescheduling and without security constraints

Performance Evaluation

- Evaluated following the guidelines of CEC2011
  - 25 independent run
  - Swarm size = 50
  - Compared with top 2 best performing algorithms
    - GA-MPC - Genetic Algorithm with a new Multi-Parent Crossover: An efficient and improved variant of GA with a new crossover operator. The algorithm has produced robust and high quality solution to optimization problems.
    - SAMODE – Self-Adaptive Multi-Operator Differential Operator: A new variant of DE with four different mutation types and one crossover operator with each operator assigned a sub-population. The algorithm has successfully addressed problems with diverse classes.

Performance Evaluation

MePSO algorithm proposed the following new lines:

$$\boxed{n_{6-2} = 3, n_{4-6} = 2, n_{3-5} = 1 \; and \; n_{1-5} = 1}$$

Total Optimum Cost (All values are x 10^3)

GA-MPC	SAMODE	MePSO
2.200E+02	2.200E+02	2.200E+02

All the algorithms have produced same results

BUT

1. MePSO has achieved the solution within 50% of function evaluations.
2. MePSO has a swarm size of 50 compared to 100 of other two algorithms
3. MePSO is computationally efficeint for real-world problems.

Publications

- Tanweer, M. R., S. Suresh, and N. Sundararajan. "Self regulating particle swarm optimization algorithm“. *Information Sciences* 294 (2015): 182-202.
- Tanweer, M. R., S. Suresh, and N. Sundararajan. Dynamic Mentoring and Self-Regulation based Particle Swarm Optimization Algorithm for solving Complex Real-world Optimization Problems. *Information Sciences*. 326 (2016): 1-24.
- Tanweer, M. R., Suresh, S., & Sundararajan, N. (2015, May). Mentoring based particle swarm optimization algorithm. *Evolutionary Computation (CEC), 2015 IEEE Congress on*. IEEE, 2015.
- Tanweer, M. R., Suresh, S., & Sundararajan, N. (2015, May). Improved SRPSO algorithm for solving computationally expensive numerical optimization problems. *Evolutionary Computation (CEC), 2015 IEEE Congress on*. IEEE, 2015.

Additional Info.

- Papers can be downloaded from ResearchGate https://www.researchgate.net/profile/Muhammad_Tanweer
- For Software: Please contact M.R. Tanweer at muhammad170@e.ntu.edu.sg

Some Open problems

- Issue in McRBF: Current framework is a static implementation of metacognition
  - Fixed control signals
  - Monitory signals are based on current samples. They do not reflect feeling-of-knowing, judgment-of-knowledge and ease-of-learning
- Human Thinking
  - Common sense influence learning significantly
  - Introspective/retrospective thinking influence the learning significantly. They are responsible for dynamic change in control from meta-cognition
- Social Learning
  - Current framework does not consider the cooperation/collaboration in learning
- Transfer Knowledge – one domain to other

THANK YOU

Diagnosis of Alzheimer's Disease (contd)

- **Neuroimaging:**
  - **Computed Tomography (CT):** O. L. Lopez et. al, Computed tomography but not magnetic resonance imaging identified periventricular white-matter lesions predict symptomatic cerebrovascular disease in probable Alzheimer's disease, Archives of Neurology, vol. 52, pp. 659-664, 1995.
  - **Single-Photon Emission Computed Tomography (SPECT):** J. Ramrez et. al, Early detection of the Alzheimer's disease combining feature selection and kernel machines. Advances in Neuro-Information Processing, vol. 5507, pp. 410-417, 2009.
  - **Positron Emission Tomography (PET):** M. Lopez et.al, Principal component analysis based techniques and supervised classification schemes for the early detection of Alzheimer's disease, Neurocomputing, vol. 74, pp. 1260-1271, 2011.
  - **Magnetic Resonance Imaging (MRI):** S. Kloppel et.al, Automatic classfication of MR scans in Alzheimer's disease, Brain, vol. 131, pp. 681-689, 2008.